PRACTICAL ANALYSIS

GRAPHICAL AND NUMERICAL METHODS

Practical Analysis

GRAPHICAL AND NUMERICAL METHODS

By Dr. Fr. A. Willers, PROFESSOR AT THE FREIBERG MINING ACADEMY

Translated by Robert T. Beyer, Ph. D., ASSISTANT PROFESSOR OF PHYSICS, BROWN UNIVERSITY

DOVER PUBLICATIONS, INC.
New York

PRINTED IN THE UNITED STATES OF AMERICA
THE WILLIAM BYRD PRESS, INC.
RICHMOND VIRGINIA

TRANSLATOR'S PREFACE

The translation of this volume follows the German edition except for Articles 3 and 6, which deal with the slide rule and calculating machines. In these cases, the material has been revised or entirely rewritten, in order to describe the equivalent instruments of American design.

The diagrams in this edition are, for the most part, reprints of those of the original volume, and may therefore contain some German symbols, e.g., the use of the comma to denote the decimal point.

The translator wishes to thank his wife, Ellen F. Beyer, for her assistance in the preparation of the manuscript for publication.

ROBERT T. BEYER

Providence, February, 1947.

FOREWORD

The following volume describes numerical, graphical and a few instrumental methods of practical analysis. Even if the numerical methods, in which the approximation can be carried as far as desired, are the most important, I still believe it necessary to describe the graphical methods, since I am of the opinion (in opposition to the views of other authors) that they are of practical importance. Certainly, their accuracy is often not very great. Yet in many cases, the approximation achieved by them is sufficient, and in other cases, the data obtained from the graphical method can be a useful starting point for numerical methods. In any case, many will prefer the graphical methods for their clarity, and their more convenient manipulation, in contrast to numerical methods which frequently require so much computational work. Of course, if it is a question of greater accuracy, numerical methods must be employed.

I have endeavored so to arrange the material that each of the six chapters is intelligible by itself, if one occasionally refers back to the earlier material, particularly to the chapter on interpolation. To keep the volume from being too large, I have selected from the extensive material available only what seemed important to me. I have regretfully omitted many topics, e.g., everything which deals with methods of mathematical statistics. Nevertheless, I hope that even the accomplished worker in the field will find something new here and there. Brevity is avoided in the presentation, in order to make the volume more intelligible to the beginner, and the material is illustrated by numerous examples.

In conclusion, I wish to express my gratitude to all those who have assisted me in the preparation of this book. Above all, I must thank Prof. R. Rothe for many suggestions which I have received from him in the course in applied mathematics at the Institute for Applied Mathematics at the Berlin Technische Hochschule, courses in which I was permitted to collaborate with him for several years. A series of examples stems from this work, as well as the observations of Art. 17 relating to the theory of the planimeter. Finally, Messrs. W. Raabe and H. J. Luckert have assisted me in the examination of the proofs; in particular the latter took the trouble of working out most of the examples.

FR. A. WILLERS.

Charlottenburg, February, 1928.

TABLE OF CONTENTS

Chapter Five

ANALYSIS OF EMPIRICAL FUNCTION

Chapter Six

APPROXIMATE INTEGRATION OF ORDINARY DIFFERENTIAL EQUATIONS

CHAPTER ONE

NUMERICAL CALCULATION AND ITS AIDS

1. Calculations with Approximate Values.

1. In the application of mathematics to practical problems, we are usually dealing with *inaccurate data*. In the first place, all measurements entering into the calculation involve errors of some magnitude. In addition, many of the numerical values used, particularly the irrational numbers, such as certain roots, e, π, etc., are rounded off. The *absolute errors* of such quantities are easily estimated. When the rounding off is done correctly, these errors amount at most to one half unit in the last place retained. Therefore, when numbers are rounded off, it makes a difference whether or not zeros are inserted after the last digit of a decimal fraction. For example, the number 0.76300 is stated with a hundred times greater accuracy than 0.763. In the first case the inexactness amounts at most to 5×10^{-6}, but in the second it can be as large as 5×10^{-4}. We are often dependent on an estimate of the inaccuracy involved in the measurements. Nevertheless, the limits of error are usually known for measurements which are carefully performed.

The *worth* of an approximate value or of a measurement cannot be judged by the size of the absolute error, but only by the *relative* or *percentage error.* If a' is found by measurement for a quantity the true value of which is a, then the *absolute error* is

$$\Delta = a' - a. \tag{1}$$

The negative quantity $-\Delta$, which must be added to the measured value to get the true value, is known as the *correction.* Relative error is written

$$\delta = \left| \frac{a' - a}{a} \right| = \left| \frac{\Delta}{a} \right| \approx \left| \frac{\Delta}{a'} \right|, \tag{2}$$

while percentage error is given by

$$\epsilon = 100\delta = 100 \left| \frac{\Delta}{a} \right|. \tag{3}$$

Both of these numbers are dimensionless.

Example: The major axis of the earth was found by Bessel (1837) to be $a' = 6{,}377{,}397$ meters, and by Helmert, $a = 6{,}378{,}200$ meters.

If Helmert's value is taken as the correct one, keeping in mind that the first result is expressed in legal meters[1],* while the second is expressed in international meters, the absolute error of Bessel's result is really considerable, namely $|\Delta| > 800$ m; on the other hand, the relative error $\delta = |\Delta|/a = 0.000126$, and also the percentage error $\epsilon = 0.0126\%$, are actually small. The results are therefore relatively precise.

On the other hand, Ladenburg (1926)[2] found as the best value for Planck's constant, $h = (6.55 \pm 0.01) \times 10^{-27}$ erg sec., while Planck (1913) gave the value 6.41×10^{-27} erg sec. Then the absolute error $|\Delta| = 14 \times 10^{-29}$ is certainly extremely small, but the relative error $\delta = 0.02$, or the percentage error $\epsilon = 2\%$, are considerable when compared to the example above.

2. The error of the result of a calculation which is a consequence of the inaccuracy of the data employed is known as an *error of data.* Since the calculation of inaccurate data is carried out only approximately, an additional inaccuracy enters into the result, the *error of calculation.* Naturally this error must not be greater than the error of data, since full use would not then have been made of the accuracy of the data. On the other hand, if the calculation error is not appreciably smaller than the error of data, then the computer only makes unnecessary work for himself, and implies a non-existent accuracy in the result. As a rule, the error of calculation should amount to about one-tenth of the error of data. In machine calculations, however, more precise computations can be made without extra work, and the result may then be rounded off to a precision corresponding to the data.

The magnitude of the error of calculation allowed determines the choice of the method of calculation to be used. This problem will be discussed later. Here we consider two questions. First, for given errors of data, *what is the maximum inaccuracy to be expected in the result?* Second, what precision must the data have in order that the error of the result does not exceed a previously determined magnitude? Usually this second question must be decided separately for each case. Here we shall be concerned chiefly with the first question, and we shall determine an upper limit for the errors which may possibly be introduced. It is not difficult in individual examples to give the mean error as calculated by the method of least squares.

3. Suppose that we have a formula which is a function of three quantities with real values x, y, z, the absolute errors of which are Δx, Δy, Δz. Thus

*Notes and references to the literature are collected at the end of each section.

measurement furnishes the values $x + \Delta x$, $y + \Delta y$, $z + \Delta z$. The *absolute error of the result* is then

$$\Delta u = f(x + \Delta x, y + \Delta y, z + \Delta z) - f(x, y, z)$$

or, if $f(x, y, z)$ has continuous derivatives,

$$\Delta u = f_x(\xi_1, \eta_1, \zeta_1)\Delta x + f_y(\xi_2, \eta_2, \zeta_2)\Delta y + f_z(\xi_3, \eta_3, \zeta_3)\Delta z,$$

where the $\xi_1 \cdots \zeta_3$ are certain values in the intervals x to $x + \Delta x$, y to $y + \Delta y$, z to $z + \Delta z$. These values are those for which the derivatives each have their maximum value for the interval considered. Then, if the absolute values of both the derivatives and the errors are used (so that the individual terms do not cancel each other), we obtain the maximum absolute error of the calculated expression:

$$(4) \qquad |\Delta u| \leqq |f_x|\,|\Delta x| + |f_y|\,|\Delta y| + |f_z|\,|\Delta z|.$$

The extension of this formula to n quantities follows immediately.

Example: To obtain the radius of curvature of a plano-convex spherical lens, we measure the diameter d and the sagitta s. Because the lens is not perfectly spherical, d can be determined only within 1 mm, while s can be measured to a precision of 0.1 mm.

Simple considerations give the radius of curvature as

$$r = \frac{d^2}{8s} + \frac{s}{2}.$$

The maximum absolute error is therefore

$$|\Delta r| \leqq \left|\frac{d}{4s}\right| |\Delta d| + \left|\frac{-d^2}{8s^2} + \frac{1}{2}\right| |\Delta s|$$

$$= \left|\frac{d}{4s}\right| \times 1 + \left|-\frac{d^2}{8s^2} + \frac{1}{2}\right| \times 0.1 \text{ mm.}$$

If s is approximately 3.9 mm., and d is 126 mm., then

$$r \approx 509 + 2 \text{ mm.} \approx 511 \text{ mm.},$$

$$|\Delta r| = 8.35 \times 1 + 138.5 \times 0.1 \approx 22.3 \text{ mm.}$$

Therefore $r = 51 \pm 2$ cm.[3]

4. With certain elementary functions, simple laws can be given for the *relative error.* From the formula for the absolute error, the maximum relative error is found to be

$$\delta = \left|\frac{\Delta}{f}\right| = \left|\frac{f_x}{f}\right| \cdot |\,\Delta x\,| + \left|\frac{f_y}{f}\right| \cdot |\,\Delta y\,| + \left|\frac{f_z}{f}\right| \cdot |\,\Delta z\,|.$$

For the case of a product of three factors, this gives

$$\begin{aligned} \delta &= \left|\frac{yz}{xyz}\right| \cdot |\,\Delta x\,| + \left|\frac{xz}{xyz}\right| \cdot |\,\Delta y\,| + \left|\frac{xy}{xyz}\right| \cdot |\,\Delta z\,| \\ &= \left|\frac{\Delta x}{x}\right| + \left|\frac{\Delta y}{y}\right| + \left|\frac{\Delta z}{z}\right| = \delta_x + \delta_y + \delta_z\,. \end{aligned} \tag{5}$$

The maximum relative error of a product is then equal to the sum of the relative errors of the individual factors.

Example: Two sides of a triangle are measured to be 4.32 ± 0.02 cm. and 5.67 ± 0.02 cm., and the included angle is 45 ± ¼°. The relative error of the area is then

$$\delta = \frac{2}{432} + \frac{2}{567} + (\cot 45°)(0.0044) \approx 0.0125.$$

It should be observed that the error of the angular measurement is given in radians. The maximum error in the area is then about ¼%. As we shall see later, the accuracy of a slide rule would have been sufficient for this calculation.

A simple error law also exists for the quotient of two numbers. It is

$$\begin{aligned} \delta &\leqq \left|\frac{+1/y}{x/y}\right| \cdot |\,\Delta x\,| + \left|\frac{-x/y^2}{x/y}\right| |\,\Delta y\,| \\ &= \left|\frac{\Delta x}{x}\right| + \left|\frac{\Delta y}{y}\right| = \delta_x + \delta_y\,. \end{aligned} \tag{6}$$

Therefore *the maximum relative error of a quotient is equal to the sum of the relative errors of the numerator and denominator.*

Example: This example[4] illustrates how such considerations can occasionally give an indication as to the best possible method for carrying out the measurements. The efficiency η of a transformer is first determined directly by the ratio of the power output a to the power input z, and second by the ratio of the power output to the

sum of the power output and the power loss v. The measurements of the input and output power may be carried out accurately to within 1% while the power loss may be determined within 20%. In this case η would have a value of about 0.95.

By the first method. $\delta\eta = \delta z + \delta a = 0.01 + 0.01 = 0.02$, i.e., 2%.

By the second, $\eta = a/(a + v)$,

$$\delta\eta = \left(\left|\frac{v}{(a + v)^2}\right| \Delta a + \left|\frac{a}{(a + v)^2}\right| \Delta v\right) \cdot \frac{a + v}{a}$$

$$= \left|\frac{v}{a + v}\right| \left(\left|\frac{\Delta a}{a}\right| + \left|\frac{\Delta v}{v}\right|\right)$$

$$= \left|\frac{v}{a + v}\right| (\delta a + \delta v) = (0.05)(0.21) = 0.0105, \text{ i.e., } 1\%.$$

Therefore, in spite of the far greater inaccuracy in the determination of v, the second method leads to a more accurate result.

Simple rules, similar to those for the relative error of a product or of a quotient, can easily be derived for relative errors of powers and roots. *The relative error of an nth power is equal to n times the relative error of the base number, while the relative error of an nth root is* $1/n$*th of the relative error of the radicand.*

5. The results of a calculation will be especially inaccurate whenever we deal with the *difference of two* nearly equal quantities which are known *only approximately*. To find the relative error in this case, the sum of the absolute errors, taken without regard to sign, is divided by the difference of the two given numbers.

Example: For the determination of the logarithmic decrement, we use the formula

$$\lambda = \frac{\log a_p - \log a_q}{q - p}$$

where a_p is the magnitude of the swing between the pth and the $(p + 1)$st maxima, and a_q that between the qth and the $(q + 1)$st maxima. If readings are made with mirror and scale, then the scale readings at various deflections might be as follows:

3rd and 4th maxima, 341.4 and 662.3; $a_3 = 320.9$,

6th and 7th maxima, 625.6 and 415.6; $a_6 = 210.0$.

If the accuracy of each reading is within ± 0.1, then $a_3 = 320.9 \pm$

0.2, $a_6 = 210.0 \pm 0.2$ and the relative errors are $\delta a_3 = 0.00062$, $\delta a_6 = 0.00095$. Therefore

$$\lambda = \frac{1}{3}\left[\log a_3(1 \pm \delta a_3) - \log a_6(1 \pm \delta a_6)\right]$$

$$\approx \frac{1}{3}(\log a_3 - \log a_6) \pm \frac{1}{3}(\delta a_3 + \delta a_6)$$

$$= 0.0614 \pm 0.00052.$$

The error then amounts to almost 1%.

6. If the final result is dependent on only one of several quantities to be measured, then the relative error will be

$$\delta = \left|\frac{f'(x)}{f(x)}\right| \cdot |\Delta x|. \tag{7}$$

The result becomes more accurate as the error Δx of the measurement becomes smaller. But the accuracy of the result is also greater if the ratio $|f'(x)/f(x)|$ becomes smaller. This sometimes gives an indication of *how measurements ought to be carried out in practice.* If $|f'(x)/f(x)|$ is to be made a minimum, then the derivative of this expression, with respect to x, is equated to zero. The resultant value of x is then computed.

Example: In the measurement of resistance by the bridge method, let R be the standard resistance, l the length of the wire along which the bridge contact is moved, x the length of the wire from one end up to the contact point (when zero current is measured by the galvanometer). Then the unknown resistance is

$$f(x) = R\frac{x}{l - x}.$$

From this it follows that

$$f'(x) = R\frac{l}{(l - x)^2}; \qquad \frac{f'(x)}{f(x)} = \frac{l}{x(l - x)}.$$

If the derivative of this expression is equated to zero, then

$$\frac{d}{dx}\left(\frac{f'(x)}{f(x)}\right) = l\frac{2x - l}{x^2(l - x)^2} = 0, \text{ i.e., } x = \frac{l}{2}.$$

An error of measurement therefore has the least effect on the result if the contact of the bridge is moved in the vicinity of the middle of the wire.

7. The second question mentioned in the beginning, namely, how precise must the data be in order that the *inaccuracy of the result does not exceed an assigned value,* will be illustrated here only by a single example.

Example: The area of a circle is to be determined to within 0.1%. How precisely must the radius, $r \approx 30.5$ cm. be measured, and how many places of π should be employed?

From $I = \pi r^2$ we get

$$0.001 = \left| \frac{\Delta I}{I} \right| = \left| \frac{2\Delta r}{r} \right| + \left| \frac{\Delta \pi}{\pi} \right| = 2\delta r + \delta\pi$$

If $\pi = 3.1$ is used, the relative error of π is already 0.01. This accuracy is certainly insufficient. For $\pi = 3.14$, $\delta\pi = 0.0005$. Then we must have

$$| \Delta r | \leqq 0.00025r \approx 0.0076 \text{ cm.}$$

This accuracy is rather difficult to obtain, since one can scarcely measure such lengths more accurately than within one-tenth of a millimeter without special methods. If another place is taken for π, i.e., 3.142, then for $\delta\pi = 0.00013$ and $| \Delta r | \leqq 0.00044r = 0.013$ cm. This accuracy of measurement can be obtained. If more places of π are introduced in the calculation, then the value of $| \Delta r |$ increases to at most 0.015 cm., which is not appreciably larger than the previous value. Therefore r ought to be measured within about 0.1 mm., and at least four places should be used for π. If additional measurements enter into the calculation, this method becomes much more complicated. We must then always check back to find with which of the measurements the accuracy can most easily be improved.

NOTES

1. TRANSLATOR'S NOTE: There is a slight difference in length between the international meter, which is maintained at Paris, and the meter actually prescribed by law in various European countries. The difference in the measurement of the major axis of the earth amounts to about two meters.

2. *Handbuch der Physik*, vol. 23 (Berlin, 1926).

3. Further examples are found in R. Rothe, *Höhere Mathematik I* (Leipzig, 1925), p. 106; C. Cranz, *Lehrbuch der Ballistik I* (Berlin, 1925), Art. 44.

4. Brion, *Leitfaden zum elektrotechnischen Praktikum* (Berlin, 1910), p. 4.

2. Representation of Functions, especially Function Scales.

1. The graphical representation of a functional relationship can have two purposes. Either it should give a clear picture of the functional dependence, or it should serve as the basis for some kind of calculation

operation to be carried out graphically with the values of the function. In the first case, the two dimensional representation in the form of a curve is generally chosen.

For each axis, a particular scale must be chosen, so that the representation for the prescribed range of both variables can be fitted on the paper at one's disposal. The scales should also be so chosen that the page is completely utilized. For example, if the independent variable is to be plotted for values from x_1 to x_2, and a surface of α mm. by β mm. is available on the paper, then the upper limits for the units along the abscissa and ordinate are

$$(1) \qquad E_x = \frac{\alpha}{x_2 - x_1} \text{ mm}, \qquad E_y = \frac{\beta}{y_2 - y_1} \text{ mm}.$$

Of course these values will not in general be taken as the units. Instead, for convenience in plotting, a round number, somewhat smaller than the value determined by (1), will be used.

Example: The values $y = (\nu/R)^{\frac{1}{2}}$ for the K series of the X-ray spectrum, where ν is the wave number in reciprocal centimeters, and $R = 109{,}737$ per cm. is the Rydberg constant, are to be represented as a function of the atomic numbers from element 11 (Na) to element 74 (W) on a sheet of 40×60 cm. For Na, $y = 8.757$, for W, $y = 66.095$. In general the values are accurate to two decimal places. If we plot the atomic numbers from 11 to 74 on the shorter side, we have as the unit

$$E_x = \frac{400}{63} \text{ mm.} = 6.35 \text{ mm.}$$

or 6 mm. when rounded off, and as the unit of the ordinate

$$E_y = \frac{600}{57.338} = 10.47 \text{ mm. or } 10 \text{ mm.}$$

To obtain a particularly accurate plotting of the coordinate values, a specially made coordinate paper may be used. Points may be plotted on or read off this paper to within 0.02 mm.[1]

2. The scale should be chosen as large as possible, so that it permits a reading with the same accuracy with which the measurements are made. But in general the *accuracy of reading* should not be much greater than the accuracy of measurement. Otherwise the accuracy obtained by the measurements is very likely misjudged. If the graph is made on good millimeter paper, it can be assumed that readings can be made accurately

within 0.1 mm. The inaccuracy of a reading would then be $s = 0.05$ mm. The maximum error of measurement must then correspond to this inaccuracy of reading. In this case then, we obtain as the unit of length, or scale modulus:

$$E = \frac{s}{\Delta} \text{ mm.} = \frac{0.05}{\Delta} \text{ mm.} \tag{2}$$

If this is greater than the unit previously determined, we can break up the scale and draw the curve in two parts.

In the example above, we do not need to consider the abscissa values, since only integers are involved. If we assume that the values of $(\nu/R)^{\frac{1}{2}}$ are accurate to two decimals, i.e., that the error of measurement is $\Delta y = 0.005$, then the value for the scale modulus is

$$E_y = \frac{0.05}{0.005} \text{ mm.} = 10 \text{ mm.},$$

which is the same as before.

3. For graphical calculations, a one-dimensional representation of the functional relation in the form of a *functional scale* is chosen. Such a scale can be derived from the graph of the function, as shown in Fig. 2. In this figure, a double scale is drawn, consisting of one scale with equal intervals, i.e., a uniform scale, and one which is non-uniform. We can find the values on the uniform scale which are the function values corresponding to the arguments which appear on the other scale. But since the functional values are not important in the application of this scale, the uniform scale is usually omitted. The distance of a division stroke, denoted by x, from the origin on the remaining scale is $E_y \cdot f(x)$, where E_y is the scale modulus of the representation. The equation which was given in connection with the choice of the ordinate modulus also holds for the determination of this modulus. If there are available tables with a sufficient number of values of the function $f(x)$, the drawing of the curve can be omitted. Then $E_y \cdot f(x)$ can be plotted directly, as is the case for the logarithmic scale. If the individual division marks which would be obtained by direct plotting lie too far apart, it is advisable to draw the curve for the purpose of interpolating the intermediate values. A more finely divided scale is then obtained.

4. If the argument x is given an increment h, then different distances between strokes on the two sides of the double scale correspond to the same *increment h*. But since these distances must be neither too large nor

too small, equal increments h cannot always be maintained for the entire length of the scale. If a distance λ between strokes is given, then h, the unit increment of the argument, must be so chosen at the point x_1 that

$$\lambda = E_v[f(x_1 + h) - f(x_1)] \approx E_v - f'(\xi)\cdot h,$$

where ξ is a value between x_1 and $x_1 + h$. But since we are considering a small interval and are using only approximate values, we can set $\xi \approx x_1$. Therefore, at the point x_1 we have

$$h = \frac{\lambda}{E_v f'(x_1)}. \tag{3}$$

Naturally the value of h used is not the exact number which is obtained from this equation, but is the nearest of the values

$$\frac{1}{10^n}, \quad \frac{2}{10^n}, \quad \frac{5}{10^n}, \quad \frac{10}{10^n}, \quad \frac{20}{10^n} \cdots . \tag{4}$$

The length of the increment is therefore repeated. Only in the jump from two to five is this duplication not kept closely within the decimal system.

To determine the increment of the argument and the intervals in which the argument is to be changed by such increments, we proceed as follows. First, we choose an upper and a lower limit for the distance between strokes. These may be $\lambda_1 = 1$ mm. and $\lambda_2 = 0.5$ mm., which are values such as are chosen on the scales of a slide rule. If $f'(x)$ decreases with increasing x, then we begin with the larger length λ_1, and choose as the increment of the argument the value of the sequence (4) which lies nearest to the value

$$h_1 = \frac{\lambda_1}{E_v f'(x_1)}.$$

This increment of the argument is to be used as far as some integral value of x which lies in the neighborhood of x_2. This number is calculated from the equation

$$E_v f'(x_2) h_1 = \lambda_2 . \tag{5}$$

The next increment of the argument in the sequence, h_2, is used from this point on. Then an argument value x_3 is determined, up to which the increment h_2 is used, etc. If $f'(x)$ increases with increasing x, then we must determine h_1 by use of the smaller interval length λ_2, and x_2 with the larger, λ_1.

Example: The *increment of the argument* is to be found for the upper logarithmic scale on the slide of the ordinary slide rule. For this, $E_v = 125$ mm., and also

$$f(x) = \log x, \qquad f'(x) = \frac{\log e}{x}.$$

Since $f'(x)$ decreases with increasing x, the initial argument increment is then

$$h_1 = \frac{x \cdot \lambda}{125(0.434)} = 0.018 \approx 0.02.$$

This value is used as far as

$$x = \frac{(0.434)(125)(0.02)}{0.5} = 2.17 \approx 2.$$

From $x_2 = 2$ on, the next value of the sequence, $h_2 = 0.05$, is employed, as far as

$$x = \frac{(0.434)(125)(0.05)}{0.5} = 5.43,$$

i.e., to $x_3 = 5$. From there on, $h_3 = 0.1$ is used, up to

$$x_4 = \frac{(0.434)(125)(0.1)}{0.5} = 10.9 \approx 10,$$

i.e., to the end of the first part of the upper scale. Further investigation is not necessary since the scale from 1 to 10 is simply repeated on the second half of this scale.[2]

5. The equation

$$h = \frac{\lambda}{E_v \cdot f'(x)} \tag{3}$$

also permits us to give the *error* involved in the reading of $f(x)$. If scales are used which have been drawn accurately, with distances between successive strokes of about 1 mm., then 0.1 mm. can still be estimated. The inaccuracy in reading will then amount to about $\Delta l = 0.05$ mm. (Fig. 1). The inaccuracy of the argument x,

$$\Delta x = \frac{0.05}{E_v \cdot f'(x)}$$

FIG. 1

is therefore inversely proportional to the scale modulus which is used, and to the derivative of the function represented.

With logarithmic scales,

$$\Delta x = \frac{0.05x}{E_y(0.434)}.$$

In this case, therefore, the relative error,

$$\delta_x = \left|\frac{\Delta x}{x}\right| = \frac{0.05}{E_y(0.434)}.$$

is constant. This is an outstanding advantage of the logarithmic scale, and is enough, aside from other advantages, to justify its frequent use. For the upper scale of the slide rule, we then have

$$\delta_x = \frac{0.05}{(125)(0.434)} = 0.001\ (0.1\%).$$

For the lower scale, which is drawn with a scale modulus twice that of the upper, $E_y = 250$ mm., the relative error is half as large:

$$\delta_x = 0.0005\ (0.05\%).$$

Of course, this holds only for a well-executed scale, and for very careful reading. In rapid reading, as is usually the case with the slide rule, the probably error must be taken to be at least twice as great.

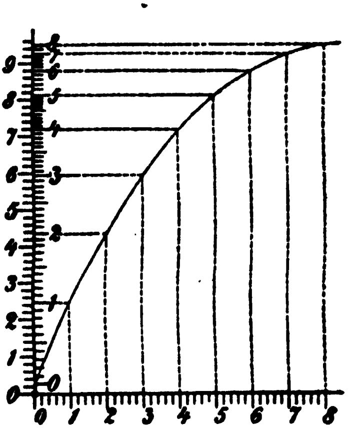


FIG. 2

6. *Function scales* can be used on both axes to replace the ordinary uniform scales in a two-dimensional representation of the functional relationship between two variables. This is done for the simplification of the graph as well as for other reasons. The functions $\xi(x)$ and $\eta(y)$ are plotted as coordinates, instead of the variables x, y themselves. In this way, for example, certain relationships can be expressed as a straight-line graph, by the use of suitable scales. The scales can also be so chosen that the graph for a particular range of the variables will be spread out, thus achieving a greater accuracy of representation in this region. Power scales and projection scales are particularly useful for such purposes. Use of logarithmic scales enables us to obtain a constant relative accuracy over the entire interval, as was shown in the preceding number. Special papers are available commercially for many of these scales (logarithmic, semilogarithmic paper).[3]

Example: A generator runs down. To ascertain the temperature at the time when the current generated becomes zero, the following temperatures T, in degrees centigrade are determined as a function of the time t by resistance measurements of the coil windings:

t	2	5	7	10	13	17
T	84.8	82.2	80.7	78.8	77.0	74.6
T-14	70.8	68.2	66.7	64.8	63.0	60.6

The cooling to the room temperature of 14° obeys the law

$$T - 14 = Ae^{at}$$

or

$$\log (T - 14) = \log A + at \log e.$$

Semi-logarithmic paper should then be used, and the values of $T - 14$ should be plotted on the logarithmic scale. The value t is plotted on the uniform scale. In the actual case the points lie very nearly

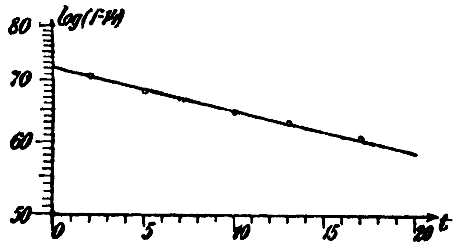


FIG. 3

on a straight line (Fig. 3). How the best possible straight line is found will be described in Art. 26. The straight line used here leads to the equation

$$T = 14 + 72.07e^{-0.00461t}$$

so that the temperature was 86.1° when the generator was shut off.

NOTES

1. Willers, *Mathematische Instrumente* (Berlin, 1926), p. 51.
2. On German slide rules, the second half of the rule is numbered from 10 to 100. See the note in Art. 3.
3. For example, hints for the use of such papers are given in Pirani, *Graphische Darstellung in Wissenschaft und Technik* (Berlin, 1922).

3. The Slide Rule[1]

1. The representation of functions by scales is used on all so-called continuous calculating devices, of which the slide rule is the best known. If two functions, $f(u)$ and $g(w)$, are represented as scales with the same scale modulus, and if these scales are set next to one another, as shown in Fig. 4, then the *fundamental equation of the slide rule* may be read from this figure:

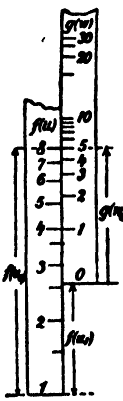


FIG. 4

$$f(u_0) = f(u_1) + g(w_0). \tag{1}$$

Calculations which can be performed with a slide rule can be derived from this equation. The ordinary slide rule possesses a number of *logarithmic scales*. On the movable portion, or *slide*, there are usually three such scales, while there are two logarithmic scales on the *stock* or fixed portion of the rule. These five scales are lettered A, B, C, CI, and D (Fig. 5 and Fig. 6). The scales A and B lie opposite each other on the upper parts of the stock and slide respectively. These scales are of the form $L \cdot \log A$, where A is the value of the function, and L, the scale modulus, is 12.5 cm. The scale is then repeated on the second half of the rule. The scales C and D lie opposite each other on the lower portion of the rule. These scales are of the form $2L \cdot \log C$ and $2L \cdot \log D$ with a scale modulus of 25 cm. The CI or inverted C scale is located in the middle of the slide. It has the same modulus as the C scale, but begins at the right side of the slide, rather than at the left.

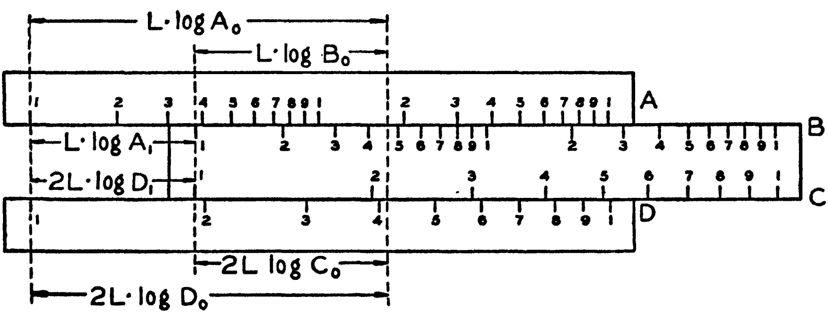


FIG. 5

The slide also usually has two trigonometric scales: S, a log sine scale of the form $2L + 2L \cdot \log \sin \varphi$ which runs from $\sin \varphi = 0.1$ to $\sin \varphi = 1$, i.e., from $\varphi = 5°44'$ to $90°$; and T, a log tangent scale of the form $2L + 2L \cdot \tan \varphi$. This scale runs from $\tan \varphi = 0.1$ to $\tan \varphi = 1$, i.e., from $\varphi = 5°44'$ to $\varphi = 45°$. The stock also has the scale L, a uniform scale of the

form $2L \cdot t$, where t varies from 0 to 1. This scale is used to find denary logarithms.

Still other scales are found on some slide rules. For example, a scale of the form $2/3L \cdot \log u$, which is repeated three times, is often included for the calculation of cube roots.

For the calculation of arbitrary powers and roots, a scale for log log v is frequently included. In practice, this scale is usually divided into three parts:

a. scale LL 3: $2L \cdot \log ln\, v$.

b. scale LL 2: $2L \cdot \log 10\, ln\, v$.

c. scale LL 1: $2L \cdot \log 100\, ln\, v$.

These scales are all set on the stock of the slide rule.

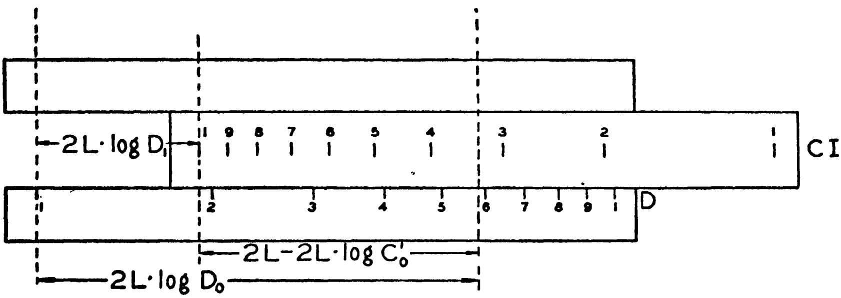


FIG. 6

2. All the scales on a slide rule *can be moved relative to each other.* This is performed either directly, by the motion of the slide with respect to the stock, or indirectly, by use of the hairline on the glass runner or indicator which can be moved along the rule. The calculations which can be carried out in this way are developed from the general slide rule equation.[2]

If we consider the A and D scales simultaneously, then, since each has the same starting point, $f(u_1) = 0$, and equation (1) becomes

$$f(u_0) = g(w_0) \tag{2}$$

$L \cdot \log A = 2L \cdot \log D = L \cdot \log D^2$. From this we have

$$A = D^2 \text{ or } D = A^{\frac{1}{2}}. \tag{3}$$

Therefore, the number on the upper scale which lies directly above a number on the D scale is the square of the latter number. Conversely, the *square root* of the number on the A scale can be read off on the D scale. It can easily be seen that the numbers with an odd number of digits

in front of the decimal are set on the first half of the A scale, and those with an even number of digits are set on the second half.

3. We now consider the *relations between scales A and B,* as shown in Fig. 5. From the figure we have

$$L \cdot \log A_0 = L \cdot \log A_1 + L \cdot \log B_0 \tag{4}$$

where A_0 and B_0 are the scale readings of any two adjacent points on the scales. In general,

$$A_1 = \frac{A_0}{B_0} = \frac{A}{B}. \tag{5}$$

If we regard the edge between the stock and the slide as a fractional bar, then all the fractions formed by adjacent pairs of numbers have the same value (a possible check on the accuracy of the scale). If a fraction A_0/B_0 is to be multiplied by a rather large sequence of numbers B_1 , B_2 , $\cdots$ B_m , then the division stroke of the slide, denoted by B_0 , is placed under the point A_0 of the A scale. The values are then read on the A scale, over the marks B_1 , B_2 , $\cdots$ B_m . Since the A and B scales contain two 1-10 intervals, the values A_1 , A_2 , $\cdots$ A_m lie opposite the values B_1 , B_2 , $\cdots$ B_m in either interval. Of course it must be kept in mind that each number is taken as a sequence of digits, without regard to the position of the decimal point. The decimal must be supplied by some estimation method. In this way, the entire set A_m can be read off from a single setting of the slide.

Example: In seven place logarithm tables, we have

$$\begin{aligned} \log \sin 42°36' &= 9.8305091 \\ \log \sin 42°35' &= 9.8303717 \\ \hline \text{Difference} &= 0.0001374. \end{aligned}$$

The following values are to be calculated:

$$\begin{aligned} \log \sin 42°35' \quad 2.5'' &= 9.8303717 + 0.0000057 = 9.8303774 \\ 18.7'' &= + 0.0000429 = 9.8304146 \\ 27.3'' &= + 0.0000626 = 9.8304343 \\ 54.2'' &= + 0.0001230 = 9.8304947. \end{aligned}$$

If we represent the number of seconds by A, then we are to form $B = (\Delta/60)\ A$. These values are then added to the value of log sin 42°35'. The last digit of the resultant number can always be in error by one unit, if the readings are not made very carefully.

To find the *decimal point* we replace the actual values by round numbers.

Example: $635(0.0083) \approx 500(0.01) = 5$, more accurately 5.27.

$7.49 \times 16.3/0.067 \approx 10 \times 20/0.1 = 2000$, more accurately 1821.

Only very rough approximations are needed, since we are interested only in the position of the decimal. If several multiplications and divisions are to be performed, a division and a multiplication can always be carried out with one setting of the slide. The intermediate results are not actually read off, but are carried along by use of the hairline.

4. The same type of calculation as discussed in (3) can also be carried out on the lower scales C and D. The accuracy of a single such reading is twice that obtained with scales A and B, because of the larger scale modulus (cf. Section 5 of Article 2, written hereafter as 2.5). However, an extra setting of the slide is often necessary with these scales. It frequently happens that the number on the C scale, to which corresponds a solution on the D scale, is beyond the entire range of the D scale. The entire length of the slide must then be pushed past the original position of one index of the slide. The original position of this index is first marked by the hairline on the glass runner, and then the slide is moved until the other index comes under the hairline. Then the solution will actually fall within the D scale interval.

Of course, time is lost in this operation, and inaccuracies enter into the calculation which can, under certain circumstances, neutralize the otherwise higher accuracy of the reading. No general rules can be given as to which scale should be used in any given operation. It should also be noted that some of the more elaborate slide rules employ additional scales to avoid this objection.[3]

5. It has been mentioned that the CI scale is *identical* in its subdivisions to the C scale, except that it begins at the opposite end of the slide (Fig. 6). If $\varphi(u)$ represents the C scale, then $2L - \varphi(u)$ represents the CI scale. From the figure,

$$2L \cdot \log D_0 = 2L \cdot \log D_1 + 2L - 2L \cdot \log C_0' \tag{6}$$

$$\log C_0'D_0 = \log D_1 + 1 = \log D_1 + \log 10$$

$$C_0'D_0 = 10D_1 = C'D \tag{7}$$

where C_0' is the scale reading on the CI scale.

If the slide is moved to the left, the expression becomes $C_0'D_0 = 10D_1$. But since this method is used only to find the digit sequence, without

regard to the decimal point, we do not need to distinguish between the two cases. With the *CI* scale, a number $D_1 = C_0'D_0$ can be divided by a series of numbers without shifting the slide.

TABLE

Slide	Stock		
	A $L \cdot \log A$	D $2L \cdot \log D$	L $2L \cdot t$
B $L \cdot \log B$	$\frac{A_0}{B_0} = A_1 = \frac{A}{B}$	$\frac{D_0}{(B_0)^{\frac{1}{2}}} = D_1$ $= \frac{D}{B^{\frac{1}{2}}}$	$\log \frac{10^{t_0}}{(B_0)^{\frac{1}{2}}} = t_1$ $= \log \frac{10^t}{B^{\frac{1}{2}}}$
C $2L \cdot \log C$	$\frac{A_0}{C_0^2} = A_1 = \frac{A}{C^2}$	$\frac{D_0}{C_0} = D_1 = \frac{D}{C}$	$\log \frac{10^{t_0}}{C_0} = t_1$ $= \log \frac{10^t}{C}$
CI $2L - 2L \cdot \log C'$	$A_0C_0'^2 = 100A_1$ $= AC'^2$	$D_0C_0' = 10D_1$ $= 10DC'$	$\log C_0'10^{t_0-1} = t_1$ $= \log C'10^{t-1}$
S $2L + 2L \log \sin \varphi$	$\frac{A_0}{\sin^2 \varphi_0} = 100A_1$ $= \frac{A}{\sin^2 \varphi}$	$\frac{D_0}{\sin \varphi_0} = 10D_1$ $= \frac{D}{\sin \varphi}$	$\log \frac{10^{t_0-1}}{\sin \varphi_0} = t_1$ $= \log \frac{10^{t-1}}{\sin \varphi}$
T $2L + 2L \log \tan \varphi$	$\frac{A_0}{\tan^2 \varphi_0} = 100A_1$ $= \frac{A}{\tan^2 \varphi}$	$\frac{D_0}{\tan \varphi_0} = \frac{10D_1}{\tan \varphi}$	$\log \frac{10^{t_0-1}}{\tan \varphi_0} = t_1$ $= \log \frac{10^{t-1}}{\tan \varphi}$

Example: The current I is to be calculated for the resistances $R_1 = 13.4$ ohms, $R_2 = 29.8$ ohms, $R_3 = 65.5$ ohms, $R_4 = 82$ ohms. The potential is 110 volts. Now $E = IR$. We use the CI scale, with the left index of the scale set over 110 on the D scale. Then, corresponding to the values of the resistance on the CI scale, the values of the current are read off the D scale:

$$I_1 = 8.22 \text{ amperes}; \; I_2 = 3.70 \text{ amperes}; \; I_3 = 1.680 \text{ amperes};$$

$$I_4 = 1.342 \text{ amperes}.$$

6. The various other relations which are listed in the accompanying table can be easily derived from the basic slide rule equation (1). In this table, it is assumed that the slide is drawn to the right. If it is drawn to the left, the value on the D scale must be multiplied by 10, while the value on the A scale must be multiplied by 100.

7. The values for $\sin \varphi$ are found with the use of the S and C scales.[4] If the hairline is set over the angle φ on the S scale, then $\sin \varphi$ can be read off the hairline on the C scale.

To find the cosine, we observe that

$$\cos \varphi = \sin (90° - \varphi). \tag{8}$$

In other words, we find the values of the $\cos \varphi$ by use of the sine of the complement of φ. For values of φ which are close to 90° (v. 4.2), the value of $\sin \varphi$ is best calculated from

$$\sin \varphi = (1 - \cos^2 \varphi)^{\frac{1}{2}} = (1 - \sin^2(90 - \varphi))^{\frac{1}{2}}. \tag{9}$$

In a similar way the values of $\tan \varphi$, $5°44' \leqq \varphi \leqq 45°$, are read off on the D scale. For values of $\varphi > 45°$, we recall that

$$\tan \varphi = \frac{1}{\cot \varphi} = \frac{1}{\tan (90 - \varphi)}. \tag{10}$$

This process can be simplified with the use of the CI scale. The values of $1/\tan (90 - \varphi)$ can then be read off directly on that scale.

The values of φ smaller than $5°44'$ are lacking on both the sine and tangent scales. To calculate these functions for the small angles, we first observe that, for small angles, both the sine and the tangent can be approximated by the angle itself, measured in radians, so that

$$\sin \varphi \approx \tan \varphi \approx \frac{\pi\varphi}{180}, \qquad \varphi < 5°44'. \tag{11}$$

The uniform L scale is used for the evaluation of logarithms. The hairline of the glass runner is set to the number on the D scale, the logarithm of which is to be found. Then the mantissa of this number appears under the hairline on the L scale.

Every trigonometric calculation affords examples for the use of the sine and tangent scales. The uniform scale may be used for the *calculation of powers of ten* with various exponents.

Example: The ratio of the strains S_2 and S_1 of the two ends of the belt of a band brake is

$$\frac{S_2}{S_1} = e^{\mu\alpha}, \tag{12}$$

where μ is the coefficient of friction, and α is the angle, in radians, through which the rod is twisted. Let $S_2 = 760$ kg., $\mu = 0.4$, $\alpha = 3\pi/2$. Then

$$S_1 = 760 \cdot e^{-0.4 \times (3/2)\pi} = 760 e^{-0.6\pi}. \tag{13}$$

First we calculate

$$\log e^{-0.6\pi} = -0.6\pi \log e = -0.6\pi(0.434) = -0.818.$$

If the slide is now lined up evenly with the stock and the hairline is moved to 0.818 on the L scale, the value of $10^{-0.818} = 0.152$ is obtained on the CI scale. A simple multiplication will then give the final result: $S_1 = 115.5$ kg.

8. A special application of the slide rule is found in the approximate solution of *quadratic equations*. This equation may be given in the form $x^2 - ax + b = 0$. From this we obtain

$$x + \frac{b}{x} = a. \tag{14}$$

where $a = x_1 + x_2$, $b = x_1x_2$. If the slide is moved until the left index of the CI scale lies over the value b on the D scale, the product of the values $C' \cdot D$, lying next to each other, is b, as is also the case with the roots of the given equation. If values of C' and D are found, by use of the hairline, for which $C' + D$ or $C' - D$ is equal to b, then these two values are the roots of the equation. Here we must pay attention to the sign of the answer as well as to the location of the decimal point. There are four cases to be considered:

I $b > 0$	a. $a > 0$	both roots positive.
	b. $a < 0$	both roots negative.
II $b < 0$	a. $a > 0$	larger root positive, smaller root negative.
	b. $a < 0$	larger root negative, smaller root positive.

Example: Let $x^2 - 4.29x - 2.623 = 0$. From the signs of a and b, it is evident that the two roots have opposite signs and that the greater, x_1, is positive. We write the equation

$$x - \frac{2.623}{x} = 4.29. \tag{15}$$

If the left index of the CI scale is set at 2.623 on the D scale, a root is then determined at about 4.8 on the D scale. For a more accurate evaluation, we now observe that $x_1 = a + |x_2|$.

Opposite $D = 4.8$ we have $C' = 0.546$. Then $\overline{D} = 4.29 + .546 = 4.836$.

Opposite $\overline{D} = 4.836$ we have $C' = 0.543$.

Then $\overline{\overline{D}} = 4.29 + .543 = 4.833$.

The limit of accuracy has been reached at this point.[5] The smaller root can be found to the third decimal place. Consequently the larger root can also be determined to the third decimal place. The roots then are

$$x_1 = 4.833, \qquad x_2 = -0.543. \tag{16}$$

In an entirely similar way, the roots of the reduced equation of the third degree may be found by use of the CI and D scales.[6]

9. The slide rule makes possible an extremely rapid calculation and is he easiest to use of all aids to calculation. It is therefore always employed henever the *accuracy* of its result is sufficient. In 2.5 it was shown nat the error in reading a slide rule can amount to 0.1% for a careful eading on an accurate scale of modulus of 12.5 cm. Now, in performing simple calculation, at least three readings are required. According to e well-known rules of errors, the resultant error then amounts of $3^{\frac{1}{2}}$ s the error of a single reading, i.e., about 0.17%. But in the actual of the slide rule, the operator does not usually take the time for such ful setting and reading, so that for rapid work, a mean error of 0.3% st be expected. When a complicated calculation is performed, in which arger number of readings is necessary, the inaccuracy of the result mes much greater. The mean error will be half as large, i.e., 0.15%

for calculations involving three readings on the lower scale of the slide rule. Slide rules used industrially with a scale modulus of 50 cm. are even more useful. The error in simple calculations on the lower scale of one of these amounts to about 0.075% or 0.1%. Recently, slide rules have been manufactured with a scale modulus of 100 cm.[7]

NOTES

TRANSLATOR'S NOTE: The material of this paragraph has been largely rewritten by the translator in order that it apply to slide rules of American design.

2. The reader should have a slide rule at hand in order to understand what follows.

3. See, for example, the Keuffel & Esser manual for the log log duplex trig slide rule (New York, 1939), pp. 12-13.

4. Some slide rules employ a different sine scale, which is then used in conjunction with the B scale. Rules for the use of such a scale can be found in most commercial manuals.

5. Further details on the method of iteration employed here will be found in 18.5.

6. For example, see Runge-König, *Numerisches Rechnen* (Berlin, 1924), p. 16 ff.

7. Other continuous calculating devices are described in Willers, *Mathematische Instrumente* (Berlin, 1926), Art. 2.

4. Linear Interpolation on Scales and in Tables

1. Up to now we have assumed that the successive strokes on th function scales lie so close together that linear interpolations can be mad by estimating within the intervals. That is, we have assumed that we car with satisfactory approximation, replace the portion of the curve lyir between two successive ordinates in the graph of the function by a straig line. But with such scales as shall be considered, this is the case only *long as the distance between the point linearly interpolated by the eye the point actually representing the value of the function remains within limit of accuracy of the reading.* For example, with a distance between strol of 1 mm., this error distance must be smaller than 0.05 mm. If we assu that the function $f(x)$ increases in the observed interval, then the distan between strokes is

$$\lambda = E_v[f(x+h) - f(x)],$$

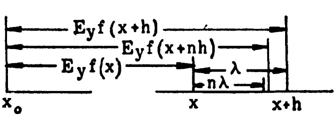


FIG. 7

and we must estimate the intermediat values

$$x + 0.1h; \qquad x + 0.2h; \qquad \cdots \qquad x + 0.9h,$$

or, more generally, $x + nh$. The estimated position of the point on the scale is therefore $l_0 = E_v f(x) + n\lambda$, while the actual position is $l = E_v f(x + nh)$. The difference between these values must be less than 0.05 mm., i.e., $|l_0 - l| \leq \Delta l \approx 0.05$ mm.

If this expression is expanded in a series which is terminated with terms of the second order, then

$$| f(x) + nf(x + h) - nf(x) - f(x + nh) | \leqq \frac{\Delta l}{E_v} \text{ or}$$

$$\left| \begin{array}{l} f(x) + nf(x) + nhf'(x) + n \frac{h^2}{2} f''(x) - nf(x) \\ -f(x) \qquad\qquad - nhf'(x) - n^2 \frac{h^2}{2} f''(x) \end{array} \right|$$

$$= \left| n(1 - n) \frac{h^2 f''(x)}{2} \right| < \frac{\Delta l}{E_v}.$$

The product $n(1 - n)$ has its maximum value for $n = 1/2$, viz., $n(1 - n) = 1/4$. If the inequality holds for $n = 1/2$, i.e., if

(1) $$\left| \frac{h^2}{8} f''(x) \right| \leqq \frac{\Delta l}{E_v},$$

then it certainly holds for the other points of the observed interval. From this it follows that if linear interpolation is to be possible, the magnitude of the argument increments h must be so chosen that if $f''(x) \neq 0$,

(2) $$h^2 \leqq \frac{8\Delta l}{E_v \mid f''(x) \mid}.$$

the magnitude of h is doubled, then the minimum distance Δl must be ur times as large, or the scale modulus must be reduced to one fourth its previous value.

2. To get some criterion for the *critical scale modulus*, i.e., the modulus r which *linear interpolation* can still be made for the usual distances between successive strokes, we assume that the argument increment h_m is used in the scale interval from x_m to x_{m+1}. If we also assume that the argument increment increases with x_m, then, according to equation 4 of Article 2 (written hereafter as 2(4)), $h_m = 2h_{m-1}$, and by 2(5), if λ is the stance between the strokes at the ends of the interval,

$$h_m = 2h_{m-1} = \frac{2\lambda}{E_v f'(x_m)}.$$

If this value for h_m is substituted in equation (2) of the preceding section, then

$$\frac{4\lambda^2}{E_v^2[f'(x_m)]^2} < \frac{8\Delta l}{E_v \mid f''(x) \mid} \qquad \text{or}$$

(3)
$$E_v > \frac{\lambda^2 \mid f''(x) \mid}{2\Delta l(f'(x_m))^2}.$$

Since Δl is approximately 0.05 mm. and $\lambda = 0.5$ mm., then

$$E_v > 2.5 \frac{\mid f''(x) \mid}{[f'(x_m)]^2} \text{ mm.}$$

Example: To use linear interpolation on a logarithmic scale without leading to incorrect results, the inequality

$$E_v > 2.5 \frac{x_m^2}{x^2 \log e} \text{ mm.}$$

must be satisfied. If we let $x = x_m$, $E_v > 5.75$ mm. For an interval with a minimum distance between strokes of 1 mm., E_v must be greater than 23 mm.

As another example, we investigate the *log sin scale* of the slide rule, for which $E_v = 12.5$ cm. The interval of the scale is sought for which a linear interpolation is no longer possible with ordinary interval lengths. Here

$$f(\varphi) = \log \sin \varphi;\ f'(\varphi) = \log e \cdot \cot \varphi;\ f''(\varphi) = -\frac{\log e}{\sin^2 \varphi}.$$

Equation (3) now becomes

$$E_v > \frac{\lambda^2 \log e}{2\Delta l \sin^2 \varphi (\log e)^2 \cot^2 \varphi_m},$$

where $\varphi \geqq \varphi_m$. If as a first approximation we set $\varphi = \varphi_m$, $\lambda = 0.5$ mm., $\Delta l = 0.05$ mm., then

$$E_v > \frac{2.5}{0.434 \cos^2 \varphi} \text{ mm.}$$

Since $E_v = 12.5$ cm., it follows that

$$\cos \varphi > \left(\frac{2.5}{0.434 \times 125}\right)^{\frac{1}{2}} = 0.215.$$

Linear interpolation therefore does not lead to incorrect results for values of φ for which $\cos \varphi < 0.215$, i.e., $\varphi > 77°$.[1]

3. In the case of *direct linear interpolation* in tables, the interpolation error $\Delta l/E_v$ mentioned in the preceding section must be smaller than one in the last place of the tabular values. If this error is denoted by Δ_t, then according to equation (1) we have

$$\Delta_\iota > \frac{h^2}{8} \mid f''(x) \mid . \tag{4}$$

For *log tables*, h can be set equal to one. Then the logarithms of the numbers 100 to 1000 are given accurately in four place tables, those of the numbers 1000 to 10000 in five place tables, and 10000 to 100000 in seven place tables. In the evaluation of the logarithm therefore, the expression

$$\frac{h^2}{8} \mid f''(x) \mid = \frac{\log e}{8x^2}$$

has a maximum at the smallest value listed in the table. For example,

Number of places in the table	4	5	6
x	100	1000	10000
Δ_ι (approximately)	$5 \cdot 10^{-6}$	$5 \cdot 10^{-8}$	$5 \cdot 10^{-10}$.

Since the errors here are materially smaller than one in the last place, linear interpolation can be used without hesitation.

Example: In *Deutscher Kalender für Elektrotechniker*, 1917, p. 26, the tangents of the angles are given, by degrees, accurate to four decimal places. In this case,

$$f(\varphi) = \tan \varphi, \qquad f'(\varphi) = 1 + \tan^2 \varphi, \; f''(\varphi) = 2 \tan \varphi(1 + \tan^2 \varphi).$$

If linear interpolation is to be possible, then $\Delta_\iota < 0.0001$, i.e.,

$$\frac{h^2}{8} \cdot 2(\tan^3 \varphi + \tan \varphi) < 0.0001,$$

or, since h will be in radians,

$$\tan^3 \varphi + \tan \varphi < \frac{0.0004}{(0.01745)^2} \approx 1.31$$

$$\varphi < 39^0 .$$

If, as is the case in the tables used in many schools, the *tangents are given for each degree to three decimal places*, linear interpolation will result in errors whenever $\varphi > 66°$. If the values of the tangents are given to only two places, errors result for $\varphi > 79°$. If such intervals do occur in tables, they should be plainly marked.

4. With *inverse interpolation*, the value of the function y is given and the corresponding argument $x + nh$ is sought. We form

$$n \cdot h = \frac{y - f(x)}{f(x + h) - f(x)} \cdot h \approx \frac{y - f(x)}{f'(x)}.$$

Neglecting terms of order higher than the second,

$$y \approx f(x) + nhf'(x) + \frac{n^2h^2}{2} f''(x) \text{ from which follows}$$

$$\frac{y - f(x)}{f'(x)} \approx nh + \frac{n^2h^2}{2} \frac{f''(x)}{f'(x)}.$$

In this case then, *linear interpolation is permissible* whenever

$$\overline{\Delta_\iota} > \frac{n^2h^2}{2} \left| \frac{f''(x)}{f'(x)} \right|, \tag{5}$$

where $\overline{\Delta_\iota}$ is the unit of the last digit interpolated. If interpolation is made from the higher or lower boundary value, according as n is larger or smaller than 1/2, the worst possible value of this expression is obtained for $n = 1/2$. Therefore we must have

$$\overline{\Delta_\iota} > \frac{h^2}{8} \left| \frac{f''(x)}{f'(x)} \right|. \tag{6}$$

For *log tables,*

$$\overline{\Delta_\iota} > \frac{1}{8x},$$

independent of the base of the logarithm. From this we obtain the following table, provided that there are no other errors involved.

Number of places in the table	4	5	7
Smallest value of x	100	1000	10000
$\overline{\Delta_\iota}$	0.00125	0.000125	0.0000125
Possible number of places in the number	5	7	9

Example: We seek the interval of the *log tan* φ *table* for which linear interpolation of the angle from the values of log tan φ is not possible. Here

$$f(\varphi) = \log \tan \varphi;\ f'(\varphi) = \frac{\log e}{\tan \varphi \cos^2 \varphi} = \frac{2 \log e}{\sin 2\varphi};$$

$$f''(\varphi) = \frac{-4 \log e \cos 2\varphi}{\sin^2 2\varphi}.$$

If linear interpolation is to be permitted, according to (6),

$$\overline{\Delta_\iota} > \frac{h^2}{4} \mid \cot (2\varphi) \mid$$

or, if h is expressed in minutes,

$$\overline{\Delta'_\iota} > \frac{h^2}{4} \mid \cot 2\varphi \mid \cdot \left(\frac{0.01745}{60}\right)^2.$$

First of all, that is certainly not true for $\varphi = 0$ or $\pi/2$. If a place table is available which gives the values in one minute steps, and if interpolation is made to seconds, the error must be smaller than $\overline{\Delta_\iota} = 0.01745/3600$, i.e.,

$$\mid \cot 2\varphi \mid < \frac{0.01745 \cdot 4 \cdot 60^2}{3600(0.01745)^2} \approx 229.$$

From this it follows that

$$2\varphi_1 > 15' \qquad \text{or} \qquad 2\varphi_2 < 179°45'.$$

Therefore, inverse linear interpolation will lead to error for values of φ for which $\varphi < 8'$ or $\varphi > 89°52'$.

In general, the *intervals in which direct and inverse linear interpolation can lead to errors* are not identical. These intervals depend on the magnitude of $f'(x)$. If $\mid f'(x) \mid > 1$, then the region of inaccuracy for direct interpolation is the greater; if $\mid f'(x) \mid < 1$, that for the inverse case is the greater. For example, in the table given above for log tan φ, direct interpolation will lead to errors if

$$\mid \cot 2\varphi \mid > 545 \mid \sin 2\varphi \mid ,$$

i.e., if $\varphi < 1°14'$ or $\varphi > 88°46'$.

5. In ordinary tables, the *error due to rounding off* the last place usually far exceeds the interpolation error. In general, the functional values appearing in such tables have been rounded off, while the argument values usually do not need to be rounded off. Therefore the error of these function values will be at most one half unit in the last place, or to $1/2\ 10^{-r}$ in r-place tables. If Δ_1 is the error of $f(x)$, Δ_2 that of $f(x + h)$, then the error of the value of the function values $f(x + nh)$ amounts, *by direct interpolation,* to

$$\Delta_a = \Delta_1 + \frac{\Delta_2 - \Delta_1}{h} \cdot nh = \Delta_1(1 - n) + \Delta_2 n.$$

In addition, there is the error Δ_3 which appears when the interpolated value of the function $f(x + nh)$ is rounded off. This will also be at most one half unit in the last place. With this included, we have

$$\Delta_a = \Delta_1(1 - n) + \Delta_2 n + \Delta_3 .$$

Since n is always positive and never larger than 1, this expression always satisfies the inequality

$$\Delta_a \leqq 10^{-r}. \tag{7}$$

Therefore the total error will be no larger than one in the last decimal place.

For *inverse interpolation*, the exact value would be

$$nh = \frac{y - f(x)}{f(x+h) - f(x)} \cdot h \approx \frac{y - f(x)}{f'(x)};$$

but the value calculated is

$$n'h = \frac{y - f(x) - \Delta_1}{f(x+h) - f(x) + (\Delta_2 - \Delta_1)} \cdot h \approx \frac{y - f(x) - \Delta_1}{hf'(x) + \Delta_2 - \Delta_1} \cdot h.$$

Since Δ_1 and Δ_2 are generally small in comparison to $y - f(x)$ and $h \cdot f'(x)$, the approximation

$$n'h \approx \frac{1}{f'(x)} (y - f(x) - \Delta_1)\left(1 - \frac{\Delta_2 - \Delta_1}{hf'(x)}\right)$$

may be made. Neglecting terms in $\Delta_1 \cdot \Delta_2$ and Δ_1^2, this becomes

$$n'h \approx \frac{1}{f'(x)}\left[y - f(x) - \Delta_1 - \frac{y - f(x)}{hf'(x)} (\Delta_2 - \Delta_1)\right].$$

Then the error due to rounding off is

$$\begin{aligned} \overline{\Delta_a} &\leqq \frac{1}{|f'(x)|} \left| (-\Delta_1 - n(\Delta_2 - \Delta_1)) \right| \\ &\leqq \frac{1}{|f'(x)|} \left| \Delta_1(1 - n) + \Delta_2 n \right| \leqq \frac{1}{2} \frac{10^{-r}}{|f'(x)|}. \end{aligned} \tag{8}$$

The error which results on rounding off the value calculated for x can be made arbitrarily small, since x can be calculated to any desired number of places.

Example: If we are to find an antilogarithm, then

$$\overline{\Delta_a} \leqq \frac{10^{-r}}{2 \log e} \cdot x \approx x10^{-r}.$$

The error of rounding off is then a maximum for the largest values of x appearing in the table. This is shown in the following table:

Number of places in the table	4	5	7
Maximum value of x	1000	10000	100000
$\bar{\Delta}_a$	0.1	0.1	0.01.

The error of rounding off will therefore be no more than a single unit in the last place ordinarily calculated from the table. The interpolation error calculated in section 4 is then negligible in comparison with these errors. Equally large errors due to rounding off are encountered in otherwise ordinary tables of functions.

6. To the errors of interpolation and rounding off should be added the error of data, discussed in Art. 1, which affects the accuracy of the number which is to be used in connection with the table. This error has the value

$$\Delta_f = |f'(x)| \cdot \Delta x. \tag{9}$$

The *total error* of a reading from a table is then

$$\Delta < |f'(x)| \cdot |\Delta x| + \frac{h^2}{8} |f''(x)| + 10^{-r}, \tag{10}$$

for *direct interpolation*. Here Δx is the inaccuracy of the initial value x, h is the difference between two successive values of x in the table, and r is the number of places in the table.

Example: In a five place log table, the log of a number which is rounded off to five digits may be obtained. In the most unfavorable case,

$$\Delta < \frac{\log e}{x} \cdot |\Delta x| + \frac{h^2}{8} \frac{\log e}{x^2} + 10^{-r}$$

$$< \frac{0.4343}{1000}(0.05) + \frac{1}{8} \times \frac{0.4343}{1000^2} + 0.00001$$

$$\approx 2 \cdot 10^{-5} + 5 \cdot 10^{-8} + 1 \cdot 10^{-5} \approx 3 \cdot 10^{-5}.$$

Therefore an inaccuracy of at most three units in the last place is to be expected.

With inverse interpolation, an upper bound can also be placed on the total possible error:

$$\bar{\Delta} < \frac{\Delta y}{|f'(x)|} + \frac{h^2}{8} \left| \frac{f''(x)}{f'(x)} \right| + \frac{1}{2} \frac{10^{-r}}{|f'(x)|}. \tag{11}$$

Example: We seek an antilogarithm from a five place log table. The error in this table is no larger than one in the last decimal place.

Then

$$\bar{\Delta} < \frac{10^{-5}x}{0.4343} + \frac{1}{8x} + \frac{1}{2}\frac{10^{-5}x}{0.4343} = \frac{3}{2}x \cdot \frac{10^{-5}}{0.4343} + \frac{1}{8x}.$$

Since, as we saw in section 4, the error of interpolation in log tables is exceedingly small, $\bar{\Delta}$ will have its maximum value for the maximum value of x, namely, 10000; therefore

$$\bar{\Delta} < \frac{3}{2} \cdot \frac{10^{-1}}{0.4343} + \frac{1}{8 \cdot 10^4} \approx 0.3.$$

The fifth digit of the number can then be in error by no more than three units.

7. Finally, let us compare the inaccuracy involved in the use of tables of addition and subtraction logarithms (cf. 19.6, as are calculated by Leonelli[2] and Gauss,[3] with the inaccuracy involved in the use of ordinary logarithmic tables. We can disregard the error of interpolation in this case. The tables of Gauss are used to go directly from log x_1 and log x_2 to log $(x_1 \pm x_2)$ without finding the numbers x_1 and x_2. In the usual present day arrangement, which is due to Wittstein,[4] for each value ξ, the value

$$\zeta = \log (1 + 10\xi)$$

is given in the tables.

In order to find log $(x_1 + x_2)$, we form

$$\xi = \log x_1 - \log x_2 = \log \frac{x_1}{x_2} = \log t$$

from the values log x_1 and log x_2, and obtain

$$\zeta = \log (1 + 10^{\xi}) = \log (1 + t) = \log \left(1 + \frac{x_1}{x_2}\right)$$

from the table. From this, we get

$$\log (x_1 + x_2) = \log x_2 + \zeta.$$

Instead of this, it is occasionally more advantageous to form $\log x_2/x_1 = \log 1/t$, and then

$$\log (x_1 + x_2) = \log x_1 + \log \left(1 + \frac{x_2}{x_1}\right) = \log x_1 + \log \left(1 + \frac{1}{t}\right).$$

Then ζ is to be added to the larger of the two logarithms.

If the log $(x_1 - x_2)$ is to be formed from the given values log x_1 and log x_2, the equation

$$\zeta = \log x_1 - \log x_2 = \log \frac{x_1}{x_2} = \log t$$

is set up. We then observe that

$$\xi = \log (10^\zeta - 1) = \log \left(\frac{x_1}{x_2} - 1\right) = \log (t - 1).$$

For a given value ζ we therefore obtain ξ from the table, just as an antilogarithm is found from ordinary log tables. Then the expression

$$\log (x_1 - x_2) = \log x_2 + \xi$$

is formed. As above, the value which is found must be added to one of the logarithms, this time the smaller one.

Also as above, we occasionally form

$$\log (x_1 - x_2) = \log x_1 + \log (t - 1) = \log x_1 + \log \left(1 - \frac{1}{t}\right),$$

where the last logarithm is always negative. In investigating the accuracy, we must keep in mind that, for Gaussian logarithms, direct interpolation is used in the first case, while inverse interpolation is used in the second.

First we shall estimate the error involved in a calculation with ordinary logarithms. Suppose log x_1 and log x_2 are free from error. Then, by equation (8), the errors of x_1 and x_2 are $(x_1 10^{-r})/(2 \log e)$ and $(x_2 10^{-r})/(2 \log e)$ respectively. The error of $x_1 \pm x_2$ is at most $((x_1 + x_2)10^{-r})/(2 \log e)$; from this it follows that, by equation (10)

log $(x_1 + x_2)$ has a maximum error $\Delta \approx 10^{-r}/2 + 10^{-r} = (3/2)10^{-r}$,
log $(x_1 - x_2)$ has a maximum error $\Delta \approx (x_1 + x_2)/2(x_1 - x_2)10^{r} + 10^{-r}$.

This latter error can be quite large for nearly equal numbers x_1 and x_2.

If, on the other hand, we use Gaussian logarithms, and if we assume, as above, that log x_1 and log x_2 are exact, then log x_1/x_2 is also exact. The error of log $(1 + x_1/x_2)$ is then, by equation (7), no larger than 10^{-r}, and since log x_2 is exact, the error of log $(x_1 + x_2)$ is also, at most, 10^{-r}.

In the calculation of log $(x_1 - x_2)$, log x_1/x_2 is again exact, by our original assumption. Since the function $f(z) = \log (1 + 10^z)$ is tabulated, then $f'(z) = (\log e \cdot 10^z \cdot \ln 10)/(10^z + 1)$ and the error of log $(x_1 - x_2)$ is, by equation (8),

$$\Delta_a = \frac{10^{-r}}{2f'(z)} = \frac{10^{-r}(1 + 10^z)}{2 \cdot 10^z} = \left|\frac{x_1}{x_1 + x_2}\right| \cdot \frac{10^{-r}}{2},$$

where it should be observed that $1 + 10^s = x_1/x_2$, so that $10^s = (x_1 - x_2)/x_2$. Here also, the maximum error is smaller than that which results from the use of ordinary logarithms.[5]

NOTES

1. These observations are extended to curved scales in Schwerdt, *Lehrbuch der Nomographie* (Berlin, 1924), Art. 16.
2. Leonelli, *Supplément logarithmique* (Bordeaux, 1802).
3. Gauss, *Zachs monatliche Korrespondenz,* 26 (1812), p. 498.
4. Wittstein, *Siebenstellige Gaussche Logarithmen* (Hannover, 1866).
5. Lüroth, *Vorlesungen über numerisches Rechnen* (Leipzig, 1900), Art. 44.

5. Nomograms.

1. *Nomograms are graphical charts in which some functional relationship of three or more variables is so represented* that it is possible, conveniently and accurately, to determine values of one variable corresponding to given values of a group of other variables. As an example, we may refer to the well-known representation of Boyle's law, $pv = c$, where p and v are chosen as coordinates. For each pair of values p and v, a value c can be read from the diagram. Also, for every p and c, a v can be found, as well as a p for each v and c. Nomograms occupy a special place among aids to calculation. For certain commonly used types of formulas, nomograms give solutions over particular ranges of the variables. They are not to be used in the same way as tables, slide rules and calculating machines. On the contrary, each nomogram is applicable for only one type of equation.

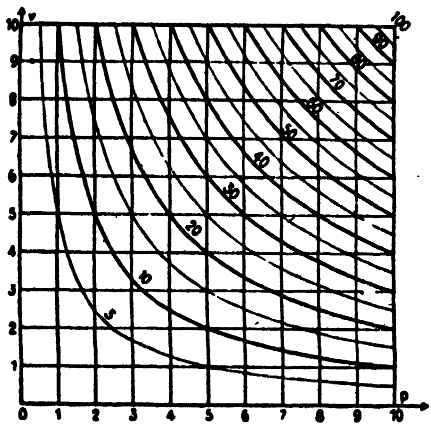


FIG. 8

The nomogram in Fig. 8 can also be used, for example, for the representation of Ohm's law, $E = IR$, or for the calculation of electric power,

$P = EI$, or for the relation of velocity, time and acceleration, $v = at$ in the case of uniformly accelerated motion, or for the lens formula, where the object and image distance are given from the focal point, $x_1x_2 = f^2$, etc. Only the letters, the units, and the particular ranges of the variables are changed. We can therefore use dimensionless numbers, only we must always keep in mind the region over which the nomogram applies. No general rules can be formulated for the transformations which must be performed on the formula to be represented in order that the resultant nomogram can be read most conveniently. These transformations must be considered in each separate case.

2. Nomograms are available commercially for a great many equations. Their preparation is often very difficult. Whether or not their manufacture is profitable depends on how difficult the nomogram is to prepare, and on how great a demand there is for the particular type. Also, the nature of the application is important in determining the type of nomogram which is chosen. For nomograms which are to be used in the laboratory, for example, we would not choose nomograms with special contrivances for reading purposes, such as a straight edge capable of rotation, in case these cannot be fastened firmly to the working table.

The reader should also note that *only approximate values* can be obtained from nomograms. The approximation, which depends on the type of the nomogram and the arrangement used, cannot be extended indefinitely, because it is limited by the size of the paper, which must not become unwieldy. The accuracy also depends on the experience and skill of the operator. He must be capable of estimating accurately to a tenth of a division, with intervals of 0.5 to 5 mm. Larger or smaller distances between division strokes or curves are not recommended. Practice is also necessary for reading non-uniform scales and curved lines.

3. *Function scales* are used for all nomograms. Besides the scales already mentioned in Art. 2, the principal function scalės are the power scales, $y = x^n$, and the projection scales, $y = ax + b/(cx + d)$. The former are used mainly because they afford an especially accurate representation in particular regions of space[1]; the latter because of their great adaptability and the ease of their construction by projection of uniform scales. Curvilinear scales are also used. Nomograms always involve the representation of a function of least three variables. We shall omit discussion of nomograms which require special reading devices. Of the other types, only charts with networks of scales, occasionally called Cartesian charts, and alignment charts, will be described. For a more detailed study, the reader is referred to the extensive literature on the subject.[2]

4. If we consider the relation $F(x, y, z) = 0$ which can in general be written in the explicit form $z = f(x, y)$, then the variables can be regarded as coordinates in space. This equation determines a surface which is most simply represented by contour lines while the xy plane is generally chosen as the plane of the drawing.

Such nomograms are especially simple to construct if the contours are straight lines or circles, and not more complicated curves, such as those of Fig. 8.

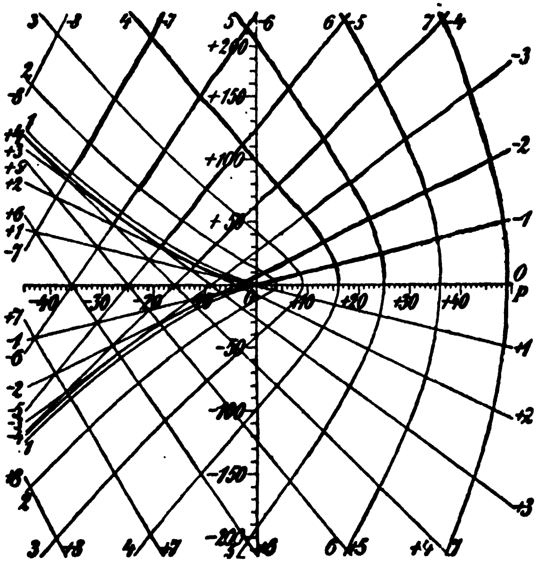


FIG. 9

Example: The nomogram reproduced in Fig. 9 can be used to obtain the roots of the reduced equation of the third degree,

$$z^3 + pz + q = 0,$$

for all real coefficients p and q. If we set $p = x$ and $q = y$, then for each real root z there is a straight line

$$\frac{x}{z^2} + \frac{y}{z^3} + 1 = 0$$

as a contour. These lines are shown in Fig. 9. We can immediately recognize (from the figure) the region over which only one real root is obtained, and for what values three real roots may be had. The two regions are separated by the discriminant curve, $q^2 = -(4/27)\ p^3$,

on which two equal real roots are obtained. A second method of treating the problem would be to draw the discriminant curve, and then lay a straight edge from the point p, q tangent to the discriminant curve. The roots could then be read from a scale at the side.[3] For example, if

$$z^3 - 37.5z - 46 = 0$$

we first locate the point with coordinates $p = -37.5$, $q = -46$. This lies between the lines numbered -1 and -2; -5 and -6; $+6$ and $+7$. An approximation yields $z_1 = -1.3$, $z_2 = -5.4$, $z_3 = +6.6$. Of course, many more lines must be drawn in, besides those shown in the figure, if the nomogram is to serve any practical purposes.

The curves which have been drawn also permit the evaluation of two complex conjugate roots. Since the coefficient of z^2 is zero, the real part of both complex roots must be equal to one half of the real root, with the opposite sign. The straight line contours may therefore be used to find the real parts of the complex roots. If we let $z = u + iv$, then the equation $z^3 + pz + q = 0$ separates into two parts,

$$u^3 - 3uv^2 + pu + q = 0,$$

$$3u^2v - v^3 + pv = 0,$$

from which we get

$$\left(\frac{v^2 - p}{3}\right)\left(\frac{2p - 8v^2}{3}\right)^2 = q^2.$$

If we again set $p = x$ and $q = y$, the equation of the curves which have been drawn are again obtained. For $v = 0$, the discriminant curve is obtained. For example, with $z^3 + 24z + 160 = 0$, we first obtain the real root $z_1 = -4$ at the point $p = 24$, $q = 160$. The real part of the complex root is then $+2$. Since the point also lies on the curve $v = +6$, the complex roots are $z_2 = 2 + 6i$, $z_3 = 2 - 6i$.

If p and q have such large values that the corresponding point no longer lies on the nomogram, or if they are so small that reading would be inaccurate, then the substitution $z = m\zeta$ may be made. The equation then becomes

$$\zeta^3 + \frac{p}{m^2}\zeta + \frac{q}{m^3} = 0$$

and m is so chosen that the point has a convenient location. For example, if $z^3 - 0.55z + 0.224 = 0$, m is set equal to 0.1.

5. Generally, each equation of the form

$$x\varphi(z) + y\psi(z) + \chi(z) = 0 \tag{1}$$

can be represented by straight lines by the use of uniform scales on the x and y axes. If the function scales $x = f(z_2)$, $y = g(z_3)$ are used, then each equation of the form

$$f(z_2)\cdot\varphi(z_1) + g(z_3)\cdot\psi(z_1) + \chi(z_1) = 0 \tag{2}$$

can be transformed into two families of straight lines, parallel to the axes, and one family of arbitrary straight lines. More generally, we can consider the three straight lines

$$a_1x + b_1y + c = 0; \qquad a_2x + b_2y + c_2 = 0; \qquad a_3x + b_3y + c_3 = 0.$$

These meet at a point if the determinant of the coefficients is zero. If now the coefficients of the first straight line are dependent only on z_1, if then $a_1 : c_1 = \varphi_1(z_1)$, $b_1 : c_1 = \psi_1(z_1)$, and if the coefficients of the second straight line depend only on z_2, etc., then three families of straight lines are obtained, the parameters of which are z_1, z_2, and z_3. Three straight lines, one belonging to each family, meet at a point if

$$\begin{vmatrix} \varphi_1(z_1) & \psi_1(z_1) & 1 \\ \varphi_2(z_2) & \psi_2(z_2) & 1 \\ \varphi_3(z_3) & \psi_3(z_3) & 1 \end{vmatrix} = 0. \tag{3}$$

Each equation which can be put in this form can be represented by three families of straight lines.

6. The transformation of a given equation into the form (3) can be effected in different ways. We naturally choose the way which yields the most accurate reading, and which is easiest to use. The equation $pv = c$ already diagrammed in Fig. 8 may be used to demonstrate how *different representations* are possible for a single equation. Here then we consider only the form of equation (2).

If we set $x = p$, $y = \lambda/v$, then $c = \lambda\, x/y$. This is a family of radiating lines which can be drawn without calculation. In France, this chart is known as a *Crépin chart* (Fig. 10). It is frequently used for multiplication. On the other hand, if we set $x = p$, $y = \lambda c$, then $v = y/\lambda x$ also forms a family of straight lines through the origin. This nomogram (Fig. 11) is known in France as *Chenevier's table*, and is used mainly for division, since, in this case, the entry lines are perpendicular to each other. Charts of this sort, in which a pencil of rays passes through a point are called *ray nomograms*. The rays can be drawn with a single straight edge which

turns on a pivot. By a projective transformation, charts can be made from these tables with *two pencils of rays.* These rays can then be replaced by two movable straight edges, the positions of which are read off on

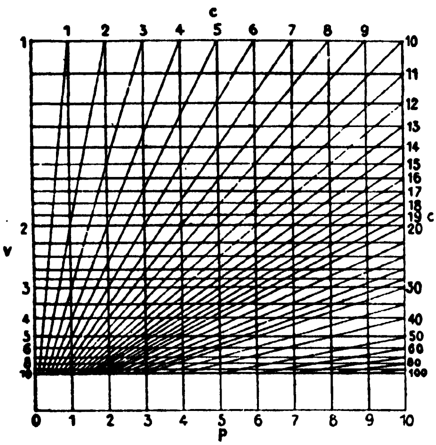


FIG. 10

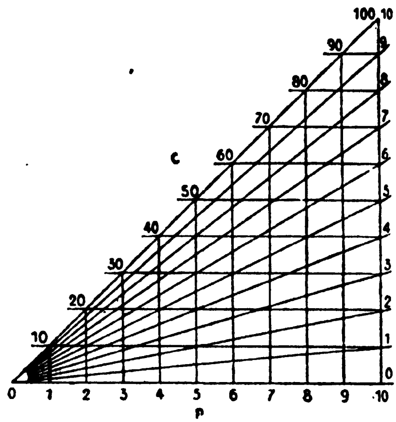


FIG. 11

scales at the side. This type of nomogram is obtained for an equation of the form $\varphi(v) \cdot \psi(p) = \chi(c)$ if function scales are used. A third possibility would be to put the equation $pv = c$ in logarithmic form: $\log p + \log v - \log c = 0$. If x is written for $\log v$ and y for $\log p$, the result is $\log c = x + y$.

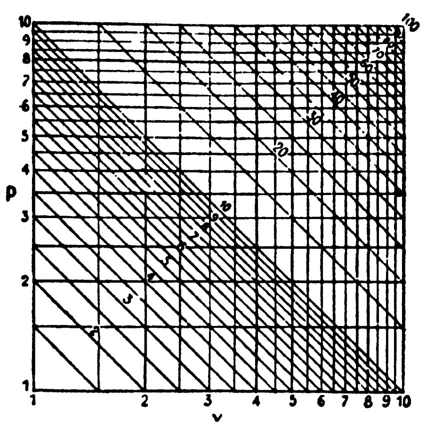


FIG. 12

D'Ocagne calls this nomogram *Lalanne's chart.* This chart is especially easy to prepare when double logarithmic paper is used.

7. If formulas are used which contain *more than three variables*, separate nomograms of three variables each must be drawn, using auxiliary variables. These nomograms are then connected by means of a family of curves as auxiliary variables (Fig. 13a and 13b). The values z_1 and z_2

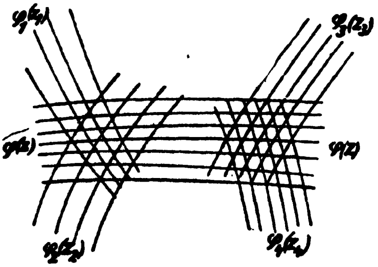


FIG. 13a

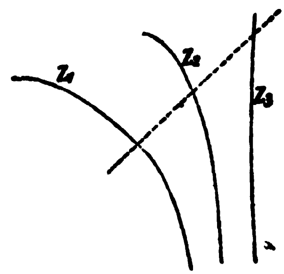


FIG. 13b

determine curves $\varphi_1(z_1)$ and $\varphi_2(z_2)$ the intersection of which determines a curve of the family $\varphi(z)$. This curve need not be numbered. We follow this curve to its intersection with the curve of the family $\varphi_3(z)$ which is determined by z_3. This in turn determines a curve of the family $\varphi_4(z_4)$, and consequently the desired value z_4 is found.

8. A second type of nomogram, the *alignment chart*, can be constructed in the following way. If a relation among the three variables is tabulated, then a scale can be prepared for each of the variables. The z_n scale would have the equation $x_n = \varphi_n(z_n)$, $y_n = \psi_n(z_n)$. The functions φ_n, ψ_n should be so chosen that the values of the three variables which satisfy the equation to be represented,

$$g(z_1, z_2, z_3) = 0, \tag{4}$$

lie on a straight line. The points are therefore *collinear*. The condition that three points be collinear is that the triangle determined by them has zero area. Therefore the determinant of the coefficients must be zero, i.e.,

$$\begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix} = \begin{vmatrix} \varphi_1(z_1) & \psi_1(z_1) & 1 \\ \varphi_2(z_2) & \psi_2(z_2) & 1 \\ \varphi_3(z_3) & \psi_3(z_3) & 1 \end{vmatrix} = 0. \tag{5}$$

Each relation (4) among the three variables which can be brought into this form can be represented by an alignment chart or by a double network of scales, since the equations (3) and (5) are identical. The question

as to whether this transformation is possible, i.e., whether the equation can be so represented, has led to detailed investigations.[4] In practical nomography, several kinds of equations have been set up which can be represented by particular types of such nomograms.

9. Of special interest are *charts with three parallel scales.* In this case, the abscissas of all the points on each scale must have the same value. Therefore,

$$\varphi_1(z_1) = c_1\ ; \qquad \varphi_2(z_2) = c_2\ ; \qquad \varphi_3(z_3) = c_3\ ,$$

i.e., equation (5) becomes

$$\text{(6)}\qquad \begin{vmatrix} c_1 & \psi_1(z_1) & 1 \\ c_2 & \psi_2(z_2) & 1 \\ c_3 & \psi_3(z_3) & 1 \end{vmatrix} = (c_3 - c_2)\psi_1(z_1) + (c_1 - c_3)\psi_2(z_2) + (c_2 - c_1)\psi_3(z_3) = 0,$$

if the determinant is expanded in the minors of the middle column. Therefore, the condition that an equation can be represented by an alignment chart with three parallel scales is that the equation can be put in the form

$$\text{(7)}\qquad C_1\psi_1(z_1) + C_2\psi_2(z_2) + C_3\psi_3(z_3) = 0,$$

where $c_1 + c_2 + c_3$ must equal zero.

Example: A nomogram is to be constructed for the bending S in mm. of a cylindrical bar of length l mm. and radius r mm., with one end clamped, and with a load P kg. on the other end. The formula for this case is

$$s = \frac{4l^3}{3\pi E r^4}\ \text{mm}^5.$$

where $S/P = s$. If we assume that the bar is of steel, then $E =$ 21,000 kg/mm.^{2}, i.e., $(E\cdot 3\pi)/4 \approx 49{,}450$. Therefore, we have

$$\log s - 3\log l + 4\log r + \log\frac{3E\pi}{4} = 0.$$

This condition on the constants can be satisfied in various ways. For example, we can write

$$-1(1 - \log s) - 3\log l + 4\left(\log r + \frac{1}{4}\log 49{,}450\right) = 0.$$

All scales are now to be drawn with the same modulus. The l and

r scales are in the same direction, while the s scale is in the opposite direction. The distances between scales are:

from the r scale to the l scale, $c_3 - c_2 = -1$
from the s scale to the r scale, $c_1 - c_2 = -3$
from the l scale to the s scale, $c_3 - c_1 = +4$

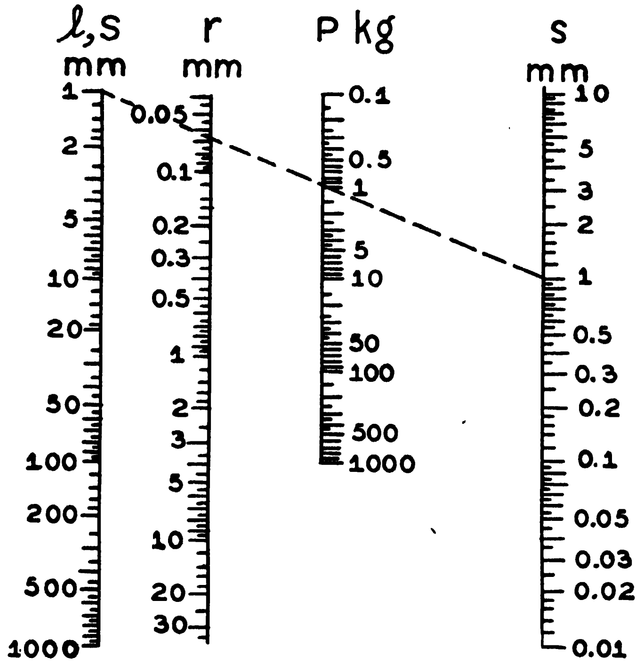


FIG. 14

as is shown in Fig. 14. The r scale is shifted somewhat, along the vertical axis. If $s = 1$ and $l = 1$ are connected by an *index line* (dotted in the figure), the point $r = 1$ is shifted about $1/4 \log 49{,}450 = 1.1735$ units toward the positive direction of the r scale. As a matter of fact, the scales can be displaced arbitrarily along these three lines. The scales are set so that the index lines cut the scales as nearly perpendicularly as possible for the most frequently used intervals of the chart. In making the drawing, any arbitrary units can be chosen. In fact, the unit for the distance between the scales can be chosen independently of the unit for the scales themselves.

A P scale is also drawn in this figure. This scale permits the determination of s, the bending per unit load, from any load P and the actual bending S. We have $S = sP$, so that

$$\log S - \log s - 2\left[\frac{1}{2}\log P\right] = 0.$$

We can use the s scale, which is already drawn. Also, the l scale can be used as an S scale. Then we need only draw the P scale, with a scale modulus one half that of the other scales, equidistant from the S and l scales, and in the same direction. Since there is no additive constant here, the points $S = 1$, $P = 1$, $s = 1$ lie on one index line.

In this case it would not be necessary to provide the line corresponding to the s scale, with an actual scale. We need only find the proper point s on this scale for the given P and S, and then connect this point with the point on the l scale which represents the length of the bar. The first reading is then made on the r scale and not on the s scale. In this method, then, we use two index lines. These serve as connections between two alignment charts and make it possible to express equations of more than three variables in this form. If a right angle or two intersecting straight edges are used to make the readings, nomograms can be prepared with four scales, on which all four variables can be read off with a single setting of the reading device.[6]

10. An equation of the form

$$(8)\qquad \begin{vmatrix} 0 & \psi_1(z_1) & 1 \\ \delta & \psi_2(z_2) & 1 \\ \varphi_3(z_3) & m\varphi_3(z_3) & 1 \end{vmatrix} = \delta(m\varphi_3(z_3) - \psi_1(z_1)) + \varphi_3(z_3)(\psi_1(z_1) - \psi_2(z_2)) = 0,$$

corresponds to a nomogram with *two parallel scales and one inclined scale*, with the equation $y = mx$. (This is not a special condition, since the origin of the coordinate system can always be put at the intersection of two scales). Here δ is the distance between the two parallel scales. From this equation, it follows that

$$(9)\qquad \psi_1(z_1)\,\frac{\varphi_3(z_3) - \delta}{\varphi_3(z_3)} - [\psi_2(z_2) - m\delta] = f(z_1)\cdot g(z_3) - h(z_2) = 0,$$

where $\varphi_3 = \delta/(1 - g(z_3))$. For example, to tabulate the equation $pv = c$ in this form, two parallel scales $f(z_1) = p$, $h(z_2) = c$ are drawn with a separation of δ. A third scale is then drawn perpendicular to these: $(m = 0)$ $\varphi_3(z_3) = \delta/(1 - v)$. If we let m depend on a fourth variable z_4, we obtain a family of scales between the two parallel scales. If we construct a curve connecting all points on these lines which have the same scale values, we get a *binary scale*, as occurs in a more general form in section

12. The nomographic type described here corresponds to an equation of the form

$$f(z_1)\cdot g(z_3) - h(z_2) + \delta k(z_4) = 0.$$

Another type of such a nomogram would be one with *three non-parallel lines* passing through a point. The corresponding equation is

$$(10)\quad \begin{vmatrix} \varphi_1(z_1) & m_1\varphi_1(z_1) & 1 \\ \varphi_2(z_2) & m_2\varphi_2(z_2) & 1 \\ \varphi_3(z_3) & m_3\varphi_3(z_3) & 1 \end{vmatrix} = \begin{vmatrix} 1 & m_1 & 1:\varphi_1(z_1) \\ 1 & m_2 & 1:\varphi_2(z_2) \\ 1 & m_3 & 1:\varphi_3(z_3) \end{vmatrix} \cdot\varphi_1(z_1)\cdot\varphi_2(z_2)\cdot\varphi_3(z_3) = 0,$$

therefore

$$\frac{m_2 - m_3}{\varphi_1(z_1)} + \frac{m_3 - m_1}{\varphi_2(z_2)} + \frac{m_1 - m_2}{\varphi_3(z_3)} = 0,$$

or

$$(11)\quad \frac{C_1}{\varphi_1(z_1)} + \frac{C_2}{\varphi_2(z_2)} + \frac{C_3}{\varphi_3(z_3)} = 0, \qquad \text{where} \qquad C_1 + C_2 + C_3 = 0.$$

This is the same form of equation as was obtained with three parallel scales, except that here the reciprocals of the functions appear. Such a nomogram would be very useful for the equation

$$\frac{1}{a} + \frac{1}{b} = \frac{2}{f}.$$

Three uniform scales would be used here, with different scale moduli. Three reciprocal scales are needed in the case of parallel scales.

The general case of three straight line scales of arbitrary position is treated most simply by the theorem of Menelaos, and leads to equations of the form

$$(12)\qquad f(z_1)\cdot g(z_2)\cdot h(z_3) = 0.$$

11. Another important type of equation can be represented by *two parallel straight line scales and one curved scale.* If one scale is placed along the y axis, the other on a line parallel to it at a distance δ, then the equation (5) becomes

$$(13)\quad \begin{vmatrix} 0 & \psi_1(z_1) & 1 \\ \delta & \psi_2(z_2) & 1 \\ \varphi_3(z_3) & \psi_3(z_3) & 1 \end{vmatrix} = \delta(\psi_3(z_3) - \psi_1(z_1)) + \varphi_3(z_3)(\psi_1(z_1) - \psi_2(z_2)) = 0.$$

From this it follows that

$$(14) \qquad \psi_1(z_1) + \psi_2(z_2)\,\frac{\varphi_3(z_3)}{\delta - \varphi_3(z_3)} + \delta\,\frac{\psi_3(z_3)}{\varphi_3(z_3) - \delta}$$

$$= f(z_1) + g(z_2)\cdot h(z_3) + k(z_3) = 0.$$

Example: We again consider the reduced equation of the third degree

$$z^3 + pz + q = 0.$$

If $f(z_1) = q_1$, $g(z_2) = p$, then $\varphi_3(z_3)$ and $\psi_3(z_3)$ must be determined by the two equations

$$z = \frac{\varphi_3}{\delta - \varphi_3}; \qquad z^3 = \frac{\delta\psi_3}{\varphi_3 - \delta}.$$

From the first of these, we get $\varphi_3 = \delta z/(1 + z) = x$. If this is substituted in the second, then

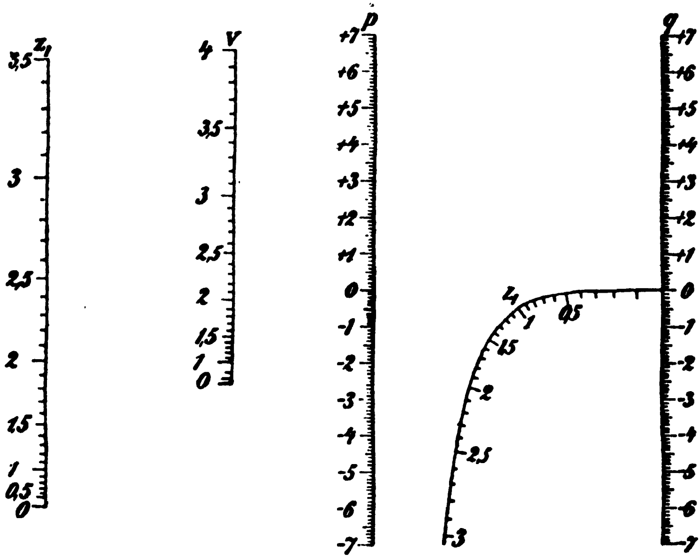


FIG. 15

$$y = \psi_3 = \frac{z^3}{\delta}\left(\frac{\delta\cdot z}{1 + z} - \delta\right) = -\frac{z^3}{1 + z}.$$

This curve is plotted in Fig. 15 for positive values of z. To get the negative roots, z is replaced by $-z$, and the positive roots of the

equation $z^3 + pz - q = 0$ are determined. The complex roots $u \pm iv$ can also be found with this nomogram. We saw in section 4 that the real part is $u = -z_1/2$, where z_1 is the real root of the equation, and that the equation

$$3u^2 - v^2 + p = \frac{3}{4}z_1^2 - v^2 + p = 0$$

is then obtained. By use of the p scale, this is represented by a nomogram with three parallel scales in the form

$$-\frac{7}{4}\left(\frac{4}{7}v^2\right) + \frac{3}{4}(z_1^2) + p = 0.$$

In this case two quadratic scales are used.

12. If a so-called *binary scale*, i.e., a network of curves dependent on two variables, is used in place of an ordinary scale, then a nomogram can be prepared with two parallel scales for an equation of the form

$$\text{(15)} \quad \begin{vmatrix} 0 & \psi_1(z_1) & 1 \\ \delta & \psi_2(z_2) & 1 \\ \varphi_3(z_3, z_4) & \psi_3(z_3, z_4) & 1 \end{vmatrix} = \delta[\psi_3(z_3, z_4) - \psi_1(z_1)] + \varphi_3(z_3, z_4)[\psi_1(z_1) - \psi_2(z_2)] = 0.$$

From (15) it follows that

$$\text{(16)} \quad \psi_1(z_1) + \psi_2(z_2)\frac{\phi_3(z_3, z_4)}{\delta - \varphi_3(z_3, z_4)} + \delta\frac{\psi_3(z_3, z_4)}{\varphi_3(z_3, z_4) - \delta} = 0.$$

A well-known example of this type of equation is the complete cubic equation

$$z^3 + nz^2 + pz + q = 0.$$

Here we put $\psi_1(z_1) = p$; $\psi_2(z_2) = q$. This gives two uniform, straight line scales. For the points of the binary scale, we must set

$$z = \frac{\varphi_3(z_3, z_4)}{\delta - \varphi_3(z_3, z_4)}; \qquad z^3 + nz^2 = \frac{\delta \cdot \psi_3(z_3, z_4)}{\varphi_3(z_3, z_4) - \delta}.$$

This gives

$$x - \varphi_3 = \frac{\delta z}{1 + z}; \qquad y = \psi_3 = -\frac{z^3 + nz^2}{1 + z}.$$

The curves $z = c$ are parallel to the p and q scales, which have uniform scales for n. The curves $n = k$ are easily constructed, starting from the curve drawn in the preceding nomogram for $n = 0$. Such a nomogram is represented in Fig. 16.

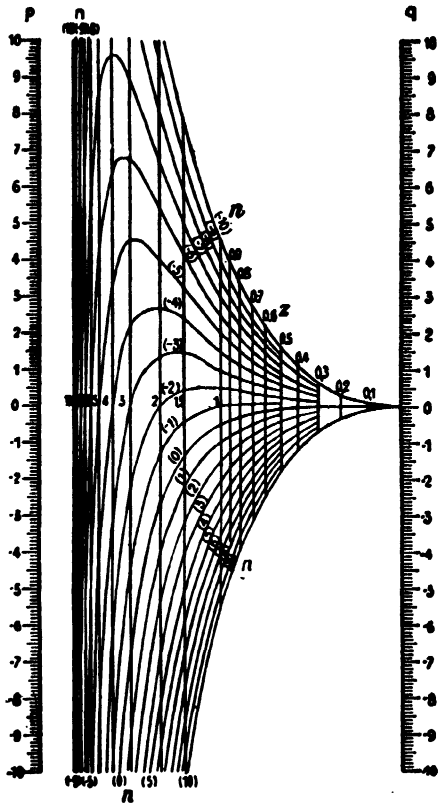


FIG. 16

NOTES

1. Schwerdt, *Lehrbuch der Nomographie* (Berlin, 1924), Art. 12.

2. For example, Massau, *Intégration graphique* (Paris-Liége, 1885); d'Ocagne, *Traité de Nomographie*, 1st ed. (Paris, 1899), 2nd ed. (1921); Sorreau, *Nomographie, un traité des Abaques* (Paris, 1921); Schwerdt, *Lehrbuch der Nomographie auf abbildungsgeometrischer Grundlage* (Berlin, 1924).

3. Schwerdt, *Zeits. f. angew. Math. u. Mech.*, IV (1924). p. 314.

4. Gronwall, *Journal de Math. pures et appl.* 8 (1912), p. 59.

5. Perry, *Angewandte Mechanik*. German edition by Schick (Leipzig, 1908), p. 143 and p. 430.

6. Cf. Dobbeler, *Zeits. f. angew. Math. u. Mech.* VII (1927), p. 485.

6. Calculating Machines[1].

1. The calculating tools previously described produce results that contain a predictable error. However, they are satisfactory for most practical

cases where the basic values are obtained by observation or measurement. The selection among these tools is determined by the requirement that the error introduced by the approximate calculation should not significantly increase the error of the result that is due to errors in the basic values.

By contrast, the accuracy of a computation on a properly designed calculating machine is limited only by the errors introduced from rounding of the right-hand digits in cases where the registers are not of sufficient capacity to show exact amounts.

The popularity of the calculating machine for mathematical work, however, rests even more on other considerations than on its inherent accuracy. It is well within the capability of the modern automatic calculator to multiply four-digit factors averaging 5's in less than four seconds, which time includes that required for setting up the amounts and for reconditioning the calculator so it may accept a similar problem. Division of similar amounts, with quotient developed to four figures averaging 5's, takes about six seconds. This time likewise includes that for figure entry and reconditioning to repeat. The calculations require a minimum of manual effort, results appear pointed off by decimal and, for many types of work, self-checking routines may be used which eliminate need of repeating the calculation to insure that no error of entry was made.

Calculating machines of the types illustrated herein which are generally used in mathematical work, perform all calculations by the rapid repeated addition or subtraction of the amount placed in a set-up mechanism or keyboard (SMA).

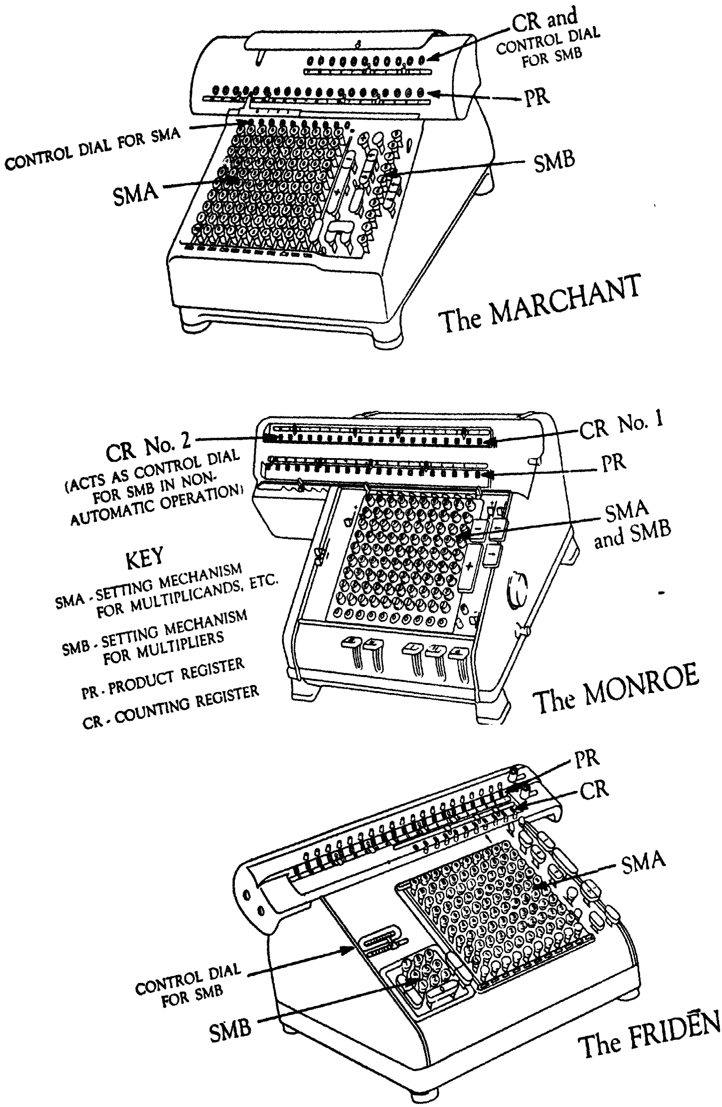
Figs. 17, 18 and 19 show three currently available American models that feature fully automatic multiplication and division. They also perform addition and subtraction and have sufficient capacity (at least 10 x 10 x 20) to enable advantage to be taken of the numerous combination techniques later to be described. The parts are lettered to correspond with the descriptive references in the text.

2. Multiplication comprises adding the amount set up in the keyboard (SMA) successively in the various carriage positions the number of times indicated by the digit of the multiplier corresponding to each carriage position. The partial products accumulate in a Product Register (PR).

Example:

512 x 37	512 is added 3 times with carriage in 2nd position	1536
	" " " 7 " " " " 1st "	3584
	Product register accumulates and shows	18944.

The number of times that the "setup" is added at each position of



Figs. 17, 18 and 19.

the carriage is shown in a Counting Register (CR). In the above case, it would show 37.

Division is the reverse process in which the dividend is first set up in the keyboard (SMA) and transferred to the Product Register (PR) usually by a touch on the Add Bar. The divisor is then set up in the keyboard (SMA), carriage is next shifted until leftmost figures of dividend and divisor are in line, and upon depression of division key the machine subtracts the divisor until the portion of the dividend directly above the divisor is reduced to an amount less than the divisor. The carriage then automatically shifts one position to the left and the process is repeated, etc.

Example:

$$512 \div 37 = 13.83\,\frac{29}{37} \qquad \begin{array}{rl} 512.00 & \\ \underline{37} & \text{Subtract 1 time} \\ 142 & \\ \underline{37} & \text{Subtract 3 times} \\ 31.0 & \\ \underline{3.7} & \text{Subtract 8 times} \\ 1.40 & \\ \underline{.37} & \text{Subtract 3 times} \\ .29 & \text{Remainder.} \end{array}$$

The counting register (CR) shows the number of subtractions in the various positions of the carriage as the desired quotient (13.83), and the Remainder (29) which appears in PR, being more than one-half of the divisor (37), indicates that the final digit should be rounded upward; the quotient is thus 13.84−, to four figures.

Modern electrically operated calculators perform the above basic computations and their combinations with a minimum of effort on the part of the operator because it is now customary for many of the clearances, shifts, tabulations, etc., to be more or less completely automatic and to require no attention by the operator.

Amounts also may be expanded in integer powers or roots obtained by suitable combinations of the basic methods; and, aided somewhat by split keyboard or product registers, different operations often may be performed at each end of the machine.

The most frequently used special procedures for mathematical work are described in a later portion of this article.

3. The machine parts and their functions are described as follows:

SMA—The principal Set-up Mechanism, in all cases a 90-key keyboard

(on 10-bank models): The Marchant (Fig. 17) has a control dial which shows figures corresponding to the setup. The Monroe (Fig. 18) facilitates verification of entry by means of a ring-shadow effect produced by the diameter of the key top being substantially less than the diameter of the opening in which it recedes.

SMB—the Set-up Mechanism for Entry of Multiplier: The Friden (Fig. 19) uses a group of ten keys with a control dial. During multiplication, the figures disappear successively from the control dial and reappear in the Counting Register (CR). The Marchant uses a single-row multiplier keyboard of ten keys. Inasmuch as it multiplies simultaneously with figure entry, the Counting Register (CR) is also the control dial. The Marchant multiplier keyboard is so designed that it will accept entry of a digit in advance of the one by which the machine is multiplying. The Monroe uses the same keyboard for entering multiplicands and multipliers; i.e., SMA is also SMB. As the multiplier is entered first, it is stored internally while the keyboard is being used for entry of multiplicand (as SMA). A recently introduced Monroe (not illustrated) has a control dial that shows the stored multiplier.

PR— The Product Register has been previously described. It shows products in multiplication, sums in addition, differences in subtraction, and the dividend and remainder in division. The Marchant and Monroe models shown in Figs. 17 and 18 provide tens carry-over in the product register to the extreme limit of the number of dials. This is effective regardless of the position of the carriage. On the Friden, entries may be made in PR by means of thumb-wheels attached to the individual dials.

CR— The Counting Register, which shows multipliers in multiplication, quotients in division, or count of items in addition or subtraction: The Friden and Marchant have single CR's with tens carry-over. The Monroe model shown has two CR's, the one at the right having tens carry-over. The left-hand CR shows individual quotients and multipliers in positive form regardless of the direction of rotation, this being accomplished by use of dials having 19 figures—nine each way from 0. The right-hand CR on the Monroe may be used to accumulate quotients or multipliers. The showing of individual quotients and the accumulated sum of a series of them is not limited, however, to calculators with double CR's, though capacity is somewhat reduced by use of alternate procedures.

There are technical differences in the models with respect to

the means of reversing the direction of rotation of CR relative to that of PR to accommodate them to the conditions in which division follows multiplication, and vice-versa. Any other reversal, such as for the purpose of registering the complement of a multiplier or quotient or of accumulating multipliers or quotients negatively, is usually done by means of a manually operated Counter Control.

CSM—Constant Storage Mechanism—found only on the Monroe model shown: It provides means for internal storage of a constant multiplier. Transfer of an amount to it is made after the said amount has previously been transferred to the internal multiplier storage referred to under the heading "SMB". The Friden and Marchant do not require internal storage of a constant because their SMA's and SMB's are separate and distinct. On them, the constant may be set up in SMA and the individual multipliers in SMB.

The use of an internal CSM enables the calculator to be used independently of the constant to perform a multiplication that does not involve the constant. However, having the constant set up as a multiplicand, as on calculators without the internal storage mechanism, instead of as a multiplier, often reduces the time of a computation because in most work involving a constant the number of multiplier digits is less than the number of digits used in the constant.

Calculators for mathematical work preferably should have tens carry-over in the dials of both PR and CR to the full extent of the registers, regardless of carriage position. Though this feature may not be necessary for single multiplications if carriage moves toward the right, it is important if products, multipliers or quotients are to be accumulated.

The manner in which the fundamental operations of multiplication, division, addition, and subtraction are performed by the various models shows many differences, which are best described by manufacturers' catalogs. Improvements that reduce the time for a complete calculating cycle "from clearance to clearance," that assure better accuracy control, or that simplify the operational procedure, are constantly being made.

4. As procedures for normal and accumulative multiplication and division, as well as for addition and subtraction, are adequately described in manufacturers' booklets, this section is limited to the description of some little known applications that may be used to advantage in mathematical work. The techniques are described in a manner that admits of their application to any of the usual machines. Though it might appear that

some of these methods would require more time than ordinary procedures, this is not the case. Each method offers either a marked reduction of time or an improvement in accuracy control by reason of elimination of copying the results of intermediate steps. Except in cases where the proof is obvious, the mathematical basis is shown. For this, the following symbols are used:

***a —The amount "a" entered (or appearing) at right of any setting mechanism or register.
a*** —Same as above, except at left.
a—Same as above, except centrally located.

In a few cases, the procedure is also exemplified by a numerical example having mathematical significance.

Instructions for the methods are presumed to be applied to calculating machines which have a capacity of not less than ten figures in CR, SMA and SMB, and of not less than twenty figures in PR.

5. *Developing formulas:* The usual procedure for solving $a \times b \times c/m \times n \times p$ etc., is, first, to evaluate the denominator by setting m in SMA, multiplying by n, then entering mn in SMA, multiplying by p, and copying mnp from PR to work sheet. Secondly, abc is similarly formed and mnp is reset in SMA from the work sheet. Depression of division key registers the desired amount in CR.

This method has the disadvantage that the group products are often so large that a single decimal setting will not suffice to carry the work to its conclusion. There is also the possibility of error when copying mnp to work sheet and re-entering in SMA.

When the result is desired to no more significant figures than may be provided by a reasonable portion of the capacity of SMA, these disadvantages may be avoided by the method of partial quotients; thus,

(1) Set up a in SMA and multiply by b; (2) set up m in SMA and divide, thus forming ab/m in CR; (3) set up c in SMA, set reverse for CR and multiply by amount appearing in CR, which reduces it to zeros; this forms abc/m in PR. (4) Set up n in SMA and divide, producing abc/mn in CR; (5) set up 1 in SMA, set reverse for CR and multiply as in step (3); this transfers the amount from CR to PR. (6) Set up p in SMA and divide, thus producing the desired result in CR.

By selecting the factors to be used in each step so that the successive amounts which develop in PR will be of approximately the same length, the result may usually be obtained to the number of figures desired without loss of precision, and within the limitations of a single setting of decimals.

6. *Three-Factor Multiplication*: The usual procedure for solving $a \times b \times c$ is well understood. However, if the sum of the number of digits in a, b, and c does not exceed the capacity of SMA, the following procedure is effective, particularly if in a series of such multiplications a is constant:

(1) Set up a at left of SMA, b at extreme right, and multiply at right of CR by c; (2) clear b from SMA at right and, by positive and negative multiplication, adjust CR so it reads same as bc that is at right of PR. PR at left shows abc.

Proof:	SMA	CR		PR forms	
1	$(a^{***}\quad {}^{***}b)$	$({}^{***}c)$	=	$^{**}ac^{**}$	$^{***}bc$
2	$(a^{***}\quad\quad)$	$^{***}(bc - c)$	=	$^{**}(abc - ac)^{**}$	
		PR accumulates		$^{**}abc^{**}$	$^{***}bc.$

A recently introduced Monroe (not illustrated) facilitates this calculation, as well as those with more factors, by use of mechanical means for transferring amounts from PR to SMA.

7. *Accumulation of Products of Three-Factor Multiplications*: A modification of the procedure just described which also permits summing the products of successive three-factor multiplications, is as follows:

(1) Set up a at right of SMA and multiply at right of CR by b.
(2) Clear SMA and CR, set up c at left of SMA, reducing the right-hand digit by the amount of 1; also set a row of 9's in SMA at right of c, to the extreme right limit of SMA.
(3) Multiply at right of CR by ab which appears directly adjacent in PR. PR at left shows abc. PR at right is thus cleared so it may contain the "ab" from the second problem of the series.

The above procedure applies only when the number of digits in a, b and c of the first set of factors does not exceed the capacity of SMA; and similarly with respect to subsequent sets, subject to further obvious limitations upon them with regard to the relation of digits to decimal.

Proof:	SMA	CR		PR forms	
1	$(\quad\quad {}^{***}a)$	$({}^{***}b)$	=		$^{***}ab$
2-3	$(c^{***}\quad {}^{**} - 1)$	$({}^{***}ab)$	=	$^{**}abc^{**}$	$^{***} - ab$
		PR accumulates		$^{**}abc^{**}$	$^{***}0$
		and so on, accumulating		$^{**}\Sigma abc^{**}.$	

8. $(a + bx)$ *When a and b are Constants and x is Variable:*

(1) Enter a in PR by means of SMA; set up b in SMA and multiply by first x which produces first value in PR; (2) without clearing, and by positive and negative multiplication, adjust CR so it reads the next x. The corresponding value of the expression appears in PR. By an analogous process $(a - bx)$ may be evaluated.

This application is extensively used in the preparation of many types of actuarial tables. See Sec. 9 for another example.

9. *$(ab + cd)$ Factors Not Exceeding* 3 *Digits Each.* This method is particularly useful when a and/or c is constant:

(1) Set up a at left of SMA and c at right; (2) multiply by d at left of CR and by b at right of CR. Central portion of PR shows desired result. An obvious space-relationship of the factors with reference to pre-set decimals is necessary.

Proof:	SMA	CR		PR forms	
1	$(a^{***}\quad {}^{***}c)$	$(d^{***}\qquad)$	$= ad^{***}$	$^{**}cd^{**}$	
1	$(a^{***}\quad {}^{***}c)$	$(\qquad {}^{***}b)$	$=$	$^{**}ab^{**}$	$^{***}bc$
		PR accumulates	ad^{***}	$^{**}(ab + cd)^{**}$	$^{***}bc.$

A frequent use of this method is for sub-tabulation, using interpolation formulas of types similar to[2]

$$f(x_n) = f(x_0) + n\delta'_{1/2} + B''_n(\delta''_0 + \delta''_1) + B'''_n\delta'''_{1/2} \tag{1}$$

in which $f(x)$, known for x_0 and x_1, is to be subtabulated at $h - 1$ intermediate values, or at the points $x_0 + n(x_1 - x_0)$ in which n comprises integer multiples of $1/h$; e.g., if the interval is to be subtabulated to tenths, n will asume values 0.1, 0.2, 0.3 $\cdots$ 0.9. Also $\delta'_{1/2}$, $(\delta''_0 + \delta''_1)$ and $\delta'''_{1/2}$ are 1st, 2nd and 3rd central differences of $f(x)$ corresponding to end values x_1 and x_0 ; the latter two differences are set up in SMA as "a" and "c" in the method above. B''_n and B'''_n are Bessel coefficients corresponding to each n; these are used as variable multipliers "b" and "d", respectively, producing the sum of the last two terms in central portion of PR for each n.

The first two terms of the above formula may be evaluated by applying the method of Sec. 8 for each n, after which the previously tabulated corrections, comprising the total of the last two terms obtained by applying the method of this section may be added (or subtracted) to produce the desired intermediate value. A continuous operation for all values of n results by applying the following:

With carriage at extreme right, $f(x_0)$ is entered at right of SMA and

added into PR. Next, $\delta'_{1/2}$ is entered in SMA in decimal relationship to $f(x_0)$ in PR, and multiplication is made by the first n, which forms the first $f(x_0) + n\delta'_{1/2}$ in PR. The corresponding previously tabulated correction-increment (total of last two terms) is then entered in SMA at its left without disturbing $\delta'_{1/2}$, which remains as a constant at right of SMA. Carriage is then shifted to left sufficiently to enable the correction-increment to be added (or subtracted), thus forming the desired value in PR. After copying, the last step is reversed, thus restoring $f_0 + n\delta'_{1/2}$ in PR. SMA at left only is then cleared of the correction increment. Carriage is then shifted to right and computation of next value is made as in Sec. 8 by building n in CR to its next higher value, and so on for all n's.

This combination of methods to facilitate sub-tabulation has obvious limitations, but it serves well when the sum of the digits in coefficients B'' and B''' does not exceed six, and similarly for the corresponding 2nd and 3rd differences. Furthermore, by use of the Comrie throw-back[3] for modifying 2nd differences, 4th differences may also be taken into account if they do not exceed 1000 in units of last place. The remaining part of the combined operation is effective if the sum of the digits in $\delta'_{1/2}$ and the correction-increment does not exceed nine. No limit is placed upon the number of digits in $f(x)$ because no more of them need be used than contained in $\delta'_{1/2}$.

10. *$(ab - cd)$ Subject to Similar Conditions as in the Preceding*:

(1) Set up a at left of SMA and c at right; (2) Reverse CR and subtractively multiply by d at left of CR; (3) disable reverse of CR and multiply by b at right of CR. Central portion of PR shows desired result. .

Proof:	SMA	CR		PR forms
1	$(a^{***} \quad {}^{***}c)$	$(-d^{***} \qquad)$	$= -ad^{***}$	$^{**} - cd^{**}$
	$(a^{***} \quad {}^{***}c)$	$(\qquad {}^{***}b)$	$=$	$^{**} + ab^{**} \quad {}^{***}bc$
		PR accumulates		$-ad^{***} \; {}^{**}(ab - cd)^{**} \; {}^{***}bc.$

11. *Evaluation of a Polynomial by Synthetic Substitution*:

The well-known synthetic-division method of obtaining the numerical value of a polynomial in x when x has a value x_1 becomes a continuous process on the calculator which upon completion shows the desired value in PR. If certain intermediate amounts are copied during the process, a second polynomial in x of one degree less is obtained, it being the quotient, omitting remainder, of the original polynomial when divided by $x - x_1$. Evaluation of this second polynomial by the same means as employe

for the first obtains the numerical value of the first derivative of the original when $x = x_1$.

The method has wide application, particularly for obtaining corrections of approximate roots of functions that are polynomials.[4]

> *Example*: Evaluate the polynomial $\cdots ex^4 + dx^3 + cx^2 + bx + a$, when $x = x_1$. (1) Set e in SMA and multiply by x, producing ex in PR; (2) change SMA to d and add, producing $ex + d$ in PR; (3) transfer this amount from PR to SMA, clear PR and multiply by x, producing $ex^2 + dx$ in PR; (4) change SMA to c and add, producing $ex^2 + dx + c$ in PR; (5) transfer this amount from PR to SMA, clear PR and multiply by x, producing $ex^3 + dx^2 + cx$ in PR; (6) change SMA to b and add, producing $ex^3 + dx^2 + cx + b$ in PR; (7) transfer this amount from PR to SMA, clear PR and multiply by x, producing $ex^4 + dx^3 + cx^2 + bx$ in PR; (8) change SMA to a and add, producing the desired value in PR.

If any term is minus, subtract instead of adding in steps 2, 4, 6. If the assumed x_1 is minus, change signs of terms having odd powers of x and multiply as if x_1 were plus. If any term in x is absent, follow method exactly but consider its coefficient as 0. If PR shows a negative amount as indicated by its being preceded by 9's, transfer it to SMA in steps 3, 5, 7 in its negative form. The product resulting from the subsequent multiplication by x_1 will be in negative form but 9's will not extend to extreme left of PR. In this case, it is advisable to have the 9's extend considerably toward the left end of PR. This may be accomplished by suitable entry of 9's in SMA, carriage shift to right, followed by adding.

If the PR readings at end of steps 2, 4, 6 are copied and preceded by e, they are the coefficients of the above-mentioned polynomial of one degree less, to which application of the method above evaluates the first derivative of the original polynomial when $x = x_1$.

This method of evaluating a polynomial and its first derivative is generally to be recommended when x_1 has several digits. If x_1 is an integer, it is easier to obtain the powers of x_1 from table and apply accumulative multiplication. The derivative in such case similarly may be evaluated from the polynomial obtained by differentiating the original polynomial.

12. *Summations of Two-Digit Paired "Scores" Obtaining* $\sum a$, $\sum b$, $\sum a^2$, $\sum b^2$, $\sum 2ab$; a frequent application in statistical mathematics.

(a) On single-keyboard calculators (SMA only) with squaring attachment:

(1) Set up first a and b at left and right of SMA, respectively, depress bar that sets amount into storage yet permits setting to remain in

SMA; (2) depress Start Key, after which a and b appear at left and right of CR. PR shows, from left to right, a^2, $2ab$, b^2. Without clearing, repeat for next pair of scores, and so on.

(b) On double-keyboard calculators (SMA and SMB):

Multiplication similar to that of the preceding requires second entry of a and b in SMB. If these amounts are read from the control dial of SMA, on models equipped with this feature, instead of from the work, a subsequent summing of $\sum a$ and $\sum b$ from the work to check $\sum a$ and $\sum b$ that appear in CR substantially proves accuracy of all calculator entries, thus avoiding need of applying other means of checking such as the Charlier Check. This same method of proof may be utilized in the case of method of 12a.

Proof:	SMA	CR		PR forms	
1	(a*** ***b)	(a***) =	a^2***	**ab**	
1	(a*** ***b)	(***b) =		**ab**	***b^2
By repeating, PR accumulates			$\sum a^2$***	**$\sum 2ab$**	***$\sum b^2$

13. *Same as above—with three-digit "scores"*:
By following the procedure outlined in Sec. 14, below, and first obtaining $\sum a$, $\sum a^2$, $\sum ab$; then repeating and obtaining $\sum b$, $\sum b^2$, and $\sum ab$, the equality of $\sum ab$ obtained from each section of the work substantially proves all entries, provided the entries into CR in all cases are made from the control dial of SMA instead of from the work.

14. *Summations of two or three-digit paired "scores" obtaining $\sum a$, $\sum a^2$,* and $\sum ab$; a frequent application in connection with obtaining a least-squares line of y on x, or x on y;

(1) Set up first b at left and corresponding a at right of SMA and multiply at right of CR by a (on calculators with Control Dial for SMA, read amount from the dial instead of from the work). (2) Repeat for next set of scores, producing $\sum ab$ at left and $\sum a^2$ at right of PR and $\sum a$ in CR. Subsequent summing of $\sum a$ from work to check $\sum a$ in CR substantially proves both entries of a. No check is had of entry of b except that provided by control dial. With three-digit scores, this process will only sum about 40 pairs, though by noting on paper the carry-over into the figures already at left of PR, a good operator may handle ten times this number of scores.
Proof:

1	(b*** ***a)	(***a)	= **ab**	***a^2
	By repeating PR accumulates		**$\sum ab$**	***$\sum a^2$.

15. *Multiplication of Amounts Which Exceed Capacity of Calculator:* A long amount, such as 345 678 901 234 567 is regarded as made up of two components 345 678 901 200 000 = a, and 34 567 = b, the whole amount being $a + b$. Similarly, the multiplier is split as $c + d$. The desired product $ac + bc + ad + bd$ has all factors within the capacity of the calculator. Often $bc + ad$ may be accumulated in PR.

16. *Evaluation of a Negative Amount:* Amounts in complementary form; i.e., which appear in PR preceded by s, may be converted to positive form by setting up the same negative amount in SMA and spacing carriage so a subtraction would produce zeros in PR. However, instead of subtracting, negatively multiply by 2, which causes the positive form of the amount to appear in PR.

Proof:	SMA	CR		
	(*** $- a$)	(*** $- 2$)	PR shows	*** $-a$
			PR forms	*** $2a$
			PR accumulates	*** a.

17. *Division of a Negative Amount:* Negative amounts in PR may be divided directly without its being necessary to convert them to positive form. On calculators with tens carry-over in CR and with provision for automatic reverse of CR upon depression of division key, proceed as follows:

(1) Set up divisor so its left-hand digit is below first significant digit in PR that is at right of 9's. (2) Multiply by any amount that will cause disappearance of 9's. (3) Manually reverse CR with respect to PR; (4) Without clearing CR, depress division key, producing the desired quotient in CR.

On models which do not have automatic reversal of CR upon depression of division key, omit step 3, or otherwise divide by subtraction. Proof: To find $| a/b |$ when given $-a$ in PR.

Let c = the integer multiplier that clears the 9's.

	SMA	CR			
				PR shows	($-a$***)
1-2	(b***)	(c***)	=	PR forms	(bc***)
		PR accumulates			[($bc - a$)****].

3-4 By division, CR forms $-(bc - a)/b$ which accumulates to the c that it already contains, producing $[c - (bc - a)/b]$ **** = $| a/b |$.

If, for example, PR shows $\cdots$ 9999432 (= −568), which it is desired to divide by 202, the process produces

$$3 - \frac{-568 + (3 \times 202)}{202} = |\,2.81188 +\,| = \left|\frac{568}{202}\right|.$$

The result is the same even if there is an over-count when clearing the 9's.

18. *Division when Dividend Exceeds Capacity of SMA and Quotient Desired to More Places than Available in CR:*

Assuming that dividend is within the capacity of PR, it is entered therein by suitable setups in SMA, shifts, and add-bar depressions. Divisor is then set up at extreme left of SMA. Division produces:

$$\frac{a}{b} = Q_{10} + R \tag{2}$$

in which Q_{10} is the quotient to ten figures (it may show only nine) and R is the remainder in PR. The calculator is cleared and R/b is then obtained to nine or ten figures. Affixing the figures of the new quotient to Q_{10} gives the desired quotient to at least 18 or 19 digits.

If a longer quotient is desired, split the dividend into parts a and b, as in Sec. 15, divide each separately by the divisor, c; extend each quotient by again dividing the remainders, as above. The desired quotient is $a/c + b/c$.

If the dividend may be contained within PR and quotient is desired only to a few more places than the capacity of CR, the process may be shortened as follows:

(1) By suitable entries in SMA and carriage shifts, transfer entire dividend to PR. (2) Set up divisor at extreme left of SMA and divide as far as CR permits. (3) Record quotient thus far and clear CR. (4) Shift carriage to right just sufficient to make room for the desired extra figures of quotient. (5) Clear SMA and reset as many of the leftmost figures of divisor as may be accommodated in SMA directly below the Remainder in PR, and divide, which produces the desired extra figures of quotient in CR.

This abbreviated method applies only if the number of digits of divisor reset in step 5 equals or exceeds the number of digits of quotient that are desired in excess of the capacity of CR.

Example: Find to 14 places 123456789098765/4567890987.

Step 2 develops quotient 27027.0874 1594437362/4567890987.

The remainder fraction is then evaluated by steps 4 and 5 as 1594437362/45679 = 34905 $\cdots$ which, when affixed gives desired

quotient as 27027.087434905 + with uncertainty in the final digit because quotient was taken to number of places equal to those of the divisor. In this case the final digit is correct.

19. *Division when Divisor Exceeds Capacity of SMA and Quotient is Desired to More Places than Available in CR*:

Applying method of Sec. 15, multiply dividend by reciprocal R of divisor N. R, to any desired precision, is obtained by applying successively

$$R_2 = R_1(2 - NR_1), \qquad \text{with error } e_2 = -Ne_1^2\,.$$

R_1 should be the reciprocal of first ten figures of N. In evaluating this iteration, only multiplication by method of Sec. 15 need be employed. Convergence is rapid.

20. When only slow-speed hand-driven calculators were available, numerous methods of solving ab/c were in use which embodied dividing simultaneously with multiplying. With modern high-speed automatic calculators, it is usually easier to form ab and then set up c and divide than it is to use these special methods for references to which see manufacturers' instructions. A similar application is that of forming two related quotients simultaneously, as below. The method is in extensive commercial use.

(a) To split an amount proportionately into two parts; i.e., $a/b \times c$ and $(b - a)/b \times c$; $a < b$., the procedure is as follows:

(1) Set up "1" at extreme right of SMA and a as close to left of SMA as possible, yet so that b may be subsequently set in SMA in the same decimal relationship as a; (2) Multiply at right of CR by c; (3) Change setup at left of SMA from a to b, clear CR and divide. This produces $(a/b) \times c$ in CR and $[(b - a)/b] \times c$ at right of PR.

The sum of the number of digits of a (or b) and c should not exceed the capacity of SMA, and the quotient CR should not be developed to more digits than capacity of SMA less digits of a (or b) and c.

Proof:	SMA		CR		PR	
1-2	$(a^{***} \quad {}^{***}1)$		$(\quad {}^{***}c)$	=	$({}^{**}ac^{**} \quad {}^{***}c)$	
	PR		SMA		CR	PR
3	$({}^{**}ac^{**} \quad {}^{***}c)$	÷	$(b^{***} \quad {}^{***}1)$	=	$(\quad {}^{***}\,ac/b)$	
					and $(\; {}^{***}c - (ac)/b)$.	

The expression at right of PR $\equiv ((b - a)c)/b$.

Subject to certain limits on capacity, this method may be adapted to cases where a exceeds b. However, because the result at right of PR

appears as a negative amount and must be so evaluated, this modification has limited practical application.

21. Iteration methods of extracting square and cube roots on calculating machines have largely superseded direct computation in cases of mathematical work. For square root, the basic application arises from the fact that dividing the number of which the square root is desired by an approximate root, and then obtaining the mean of divisor and quotient, gives a second approximation of the root substantially to twice as many figures as were correct in the first approximation; that is to say,

Let N = Number whose square root is desired
and $N^{1/2} + \epsilon$ = First approximation of square root,
then, $N/(N^{1/2} + \epsilon) = N^{1/2} - \epsilon + \epsilon^2/N^{1/2} \cdots$,

and the mean of divisor and quotient is $N^{1/2} + \epsilon^2/2N^{1/2} \cdots$ which, if taken as a second approximation of the root, has error of second order as compared with that of the first approximation.

Because ϵ^2 is always positive, the second approximation will tend to be too great numerically, so that if the first approximation contains n figures and the $2n$th figure of the quotient is odd, it should be diminished by 1 before dividing by 2.[5]

If the first approximate root is obtained from a good 25 cm slide-rule, the combined error of setting the runner on Scale A and reading from it to Scale D is generally such that after applying the division process of the calculator, the resulting second approximation of the root will be correct to five figures if the root lies in the upper portion of the rule, or to six figures if it lies in the central or lower portions.

Simplification of finding mean of divisor and quotient, is provided by use of Table (1), due to H. T. Avery, which shows for selected arguments the significant figures of double square-roots for each argument and for an argument ten times greater. Arguments are so chosen that the approximate root found by use of Table is in error not exceeding 5 in sixth place, thus giving "five-place accuracy," as ordinarily defined. The five-figure root may then readily be converted to one of ten figures by applying the mean-of-quotient-and-divisor process, as previously described.

Further refinement may be made by repeating the process, or more simply by determining the error of the ten-figure root; e.g., if R is the ten-figure root, $(= N^{1/2} + \epsilon)$, its error, ϵ, is closely

$$\epsilon = \frac{R^2 - N}{2R}; \qquad \text{and} \qquad N^{\frac{1}{2}} = R - \epsilon$$

(a) The following is the method to be used in finding five-figure square root of N from Table (1), page 69:

(1) Set up the significant figures of N at left of SMA and, with CR shifted to permit at least six quotient figures, transfer N to PR; (2) In SMA, directly below the leftmost figures of N that appear in PR, set up from Col. A of Table (1) the number nearest to the three leftmost figures of N and add into PR, which then approximates $2N$; (3) Divide this sum by the amount in Col. 1 or Col. 2 opposite the number selected from Col. A, using Col. 1 if there is an odd number of digits before the decimal point in N (or an odd number of zeros immediately following the decimal), and using Col. 2 if an even number. These divisors are the significant figures of twice the square roots of the numbers in Col. A.

The root appears in CR as quotient, there being one digit of root each way from decimal for each pair of digits in the number of which root is desired.

Examples:

Find the square root of 5331.172.

Point off in pairs, thus: '53'31'17'20, so divisor is selected from Col. 2.

$$\text{The root is} \quad \begin{array}{r} 5331172 \\ \underline{530} \\ 1456022\overline{)10631172} \\ \end{array}$$

$$\text{equals} \quad 73015\cdots$$

which is pointed off as 73.015.

Find the square root of .0005331172.

Point off into pairs, thus: .00'05'33'11'72, so divisor is selected from Col. 1.

$$\text{The root is} \quad \begin{array}{r} 5331172 \\ \underline{530} \\ 4604346\overline{)10631172} \\ \end{array}$$

$$\text{equals} \quad 23089\ldots$$

which is pointed off as .023089.

Proof: Let N = Number of which square root is desired and N is between values 10.0 and 99.9, or multiples thereof by 100. This range controls the construction of the table within the given assumptions.

δ = Difference between the leftmost three digits of N and Nearest Number from Col. A when the latter is taken in the sense that it represents the range of numbers from 10.0 to 99.9; i.e., if N = 27.8327, then from Col. A, $\delta = 27.9 - 27.8327 = 0.0673$.

ϵ = Error of the square root for 5-place accuracy, as defined; i.e.,

$< \pm 0.00005$ for N's in the range 10.0 to 99.9, the 6-place roots of which will be from $10^{1/2} \pm \epsilon$ to $10 \pm \epsilon$.

Then, by the process

(4) $$\frac{2N + \delta}{2(N + \delta)^{\frac{1}{2}}} = N^{\frac{1}{2}} + \epsilon. \qquad (\epsilon \text{ being less than 5 in 6th place.})$$

From which, neglecting minor terms

(5) $\delta^2 = 8\,N^{3/2}\epsilon$, or when $\epsilon = 0.00005$, $\delta^2 < 0.0004\,N^{3/2}$, or $\delta < 0.02\,N^{3/4}$.

The total interval between N's in Col. A, considered in the sense that they represent the range 10.0 to 99.9, is 2δ. The table is built upon this basis with some leeway to allow for the effect of the minor terms and to reduce uncertainty in the last digit of the 5-figure root because of rounding.

Example: In the vicinity of $N = 81$, δ should be $<0.02 \times 81^{3/4} = <0.54$ and the tabular interval <1.08. Col. A in the vicinity of 810, shows tabular interval of 10, which, being less than 10.8, satisfies the condition.

22. Cube root may be extracted similarly to square root when an approximate root is known. If N is the number of which cube root is desired, and $N^{1/3} + \epsilon_n = R_n$ an approximate root correct to n figures, then a root substantially correct to $2n$ figures, R_{2n}, is obtained by

(6) $$R_{2n} = \frac{1}{3}\left(\frac{N}{R_n^2} + 2R_n\right); \qquad \text{or} \qquad \frac{N + 2R_n^3}{3R_n^2}.$$

Table 2, page 71 facilitates this computation when obtaining cube roots with error not exceeding 5 in sixth place; i.e., for 5-place accuracy as ordinarily defined. This table, due to H. T. Avery, shows the significant figures of $3N^{2/3}$ when N is taken from Col. A as the amounts shown or as 1/10 or 1/100 thereof. The 5-figure root thus obtained may be extended to one of ten figures by applying the above formula for R_{2n}. If R is a ten-figure root ($=N^{1/3} + \epsilon$), its error ϵ is closely

$$\epsilon = \frac{R^3 - N}{3R^2}; \qquad \text{and} \qquad N^{1/3} = R - \epsilon.$$

23. *Finding Five-Figure Cube Root of N from Table 2, page 71*:

(1) Set up the significant figures of N at left of SMA and, with CR shifted to permit at least six quotient figures, transfer N to PR; (2) In SMA, directly below the leftmost figures of N that appear in PR, set up from Col. A of Table (2) the number nearest to the

three leftmost figures of N and multiply by 2, which causes PR to show an amount approximating $3N$; (3) Divide this sum by the factor in Cols. 1, 2 or 3 opposite the number selected from Col. A, using Col. 1 if, when N is pointed off in groups of three each way from decimal, there is one figure in the leftmost group. Similarly, select divisor from Col. 2 if the leftmost group contains two figures, or from Col. 3 if it contains three figures. These divisor factors are the significant figures of $3N^{2/3}$ when N is taken from Col. A in the sense of being equal to, 1/10, or 1/100 thereof, for Cols. 3, 2, and 1, respectively.

The root appears in CR as quotient, there being one digit of root each way from decimal for each group of three digits in the number of which root is desired.

Example: Find cube root of 5331.172.

Point off in groups of three, thus: 5′331.172′, so divisor is selected from Col. 1.

$$\begin{array}{rr} \text{The root is} & 5331172 \\ & 535 \\ & \underline{535} \\ 917678/ & \underline{16031172} \\ \text{equals} & 17469\cdots \end{array}$$

which is pointed off as 17.469.

Example: Find cube root of .0005331172.

Point off in groups of three, thus: .000′533′117′2, so divisor is selected from Col. 3.

$$\begin{array}{rr} \text{The root is} & 5331172 \\ & 535 \\ & \underline{535} \\ 1977078/ & \underline{16031172} \\ \text{equals} & 81085\cdots \end{array}$$

which is pointed off as 0.081085.

Proof: In a similar manner and with nomenclature similar to that for proving the Square Root Table (see Sec. 21), the process obtains

$$\frac{3N + 2\delta}{3(N + \delta)^{2/3}} = N^{1/3} + \epsilon. \tag{7}$$

Expanding and dropping terms containing ϵ^2 and higher, we have

$$\delta^2 = 9N^{5/3}\epsilon. \tag{8}$$

As the construction of the table is controlled by the range of N's from 100 to 1000, and multiples thereof by 1000, N may be taken as the three-figure numbers in Col. A. $N^{1/3}$ will range from 4.6− to 9.9+, then, $\epsilon < 0.00005$ if error is not to exceed 5 in sixth place. Substituting, we have $\delta < 0.0212N^{5/6}$.

Interval between successive N's $= 2\delta$, whence $2\delta < 0.042N^{5/6}$.

Example: In the vicinity of $N = 729$, $2\delta < 0.042 \times 729^{5/6} = 10.2$. It is seen that the tabular interval containing 729 is 9, and that of the next higher interval is 10.

24. The procedures described above may be adapted to higher roots. Tables of divisors (or their reciprocals as multipliers) are not difficult to compute for use in cases where there is a considerable quantity of such root extraction. C. S. Larkey has computed such divisors for fifth root.[6] For occasional extraction of higher roots, however, the logarithmic method is believed to be more suitable.

25. Calculators which multiply by the holding down of a multiplying bar a sufficient length of time to cause the desired figure to appear in CR, as well as Marchant calculators that have automatic multiplication, may multiply positively in one portion of CR and negatively in another. This may also be accomplished on the Friden and Monroe calculators with automatic multiplication by performing either the positive or negative multiplication automatically and the opposite operation by use of the multiplying bar. The method is exemplified in the following, where the problem is to find common logarithm of decimal fraction raised to a fractional power:

Example: Find $\log_{10} .9754^{.285714}$ ($\log_{10} .9754 = \bar{1}.9891828$).
(1) Set up exponent (.285714) in SMA, and in "unit" position of CR subtractively multiply by absolute value of the characteristic (−1); and at right of decimal in CR multiply by the mantissa (.9891828). This produces at right of decimal in PR the mantissa of the desired logarithm (.9969094). The characteristic of the desired logarithm is the complement of the amount at left of decimal of PR. In this case $\cdots 999$ is interpreted as $\bar{1}$.

26. The development of balancing techniques is due principally to L. J. Comrie.[7] They have greatly simplified such problems as inverse curvilinear interpolation, etc.

To find x in the equation $a + bx + f(x) \cdot c = m$ in which $f(x)$ is known for any value of x (usually from table or curve),

(1) Set up a in SMA and add into PR; (2) Set up b in SMA and multiply until PR is close to m, showing in CR a first approximate x_1, as no account has been taken of $f(x) \cdot c$; (3) From the known value of $f(x_1)$, perhaps on another calculator, separately compute $m - f(x_1) \cdot c$; (4) By direct or subtractive multiplication, adjust CR until PR is close to $m - f(x_1) \cdot c$ which was obtained in step 3, thus showing second approximate x_2 in CR; (5) In a manner similar to step 3, separately compute $m - f(x_2) \cdot c$ and continue as above until CR shows x to the number of places desired and PR shows $m - f(x) \cdot c$. In practice, the first approximations usually need be made only to a few places.

It will be observed that this method "balances" $a + bx$ and $m - f(x) \cdot c$, in conformity with the equality $a + bx = m - f(x) \cdot c$.

The following example illustrates the application of the calculator to this method, as well as its extension to cover an additional function $g(x)$, using two calculators, each developing a single side of the equation.

Example: In the following table with central differences, find x to five figures when $y = 0.38675$. This is a common problem in inverse curvilinear interpolation.

x	y				
2.0	$0.34202(f_0)$		$-1039(\delta_0'')$		$31(\delta_0^{(4)})$
		$15798(\delta_{\frac{1}{2}}')$		$-480(\delta_{\frac{1}{2}}''')$	
3.0	$0.50000(f_1)$		$-1519(\delta_1'')$		$45(\delta_1^{(4)})$.

By the Comrie throw-back, the influence of the 4th differences is taken into account by modifying the 2nd differences, as below:

(9) $M_0'' + M_1'' = (\delta_0'' + \delta_1'') - 0.184(\delta_0^{(4)} + \delta_1^{(4)}) = -0.02572.$

The Bessel interpolation formula applying is:

(10) $$f_0 + n\delta_{1/2}' + B_n''(M_0'' + M_1'') + B_n'''\delta_{1/2}''' = y_n$$

which is in the form:

$a + xb + f(x) \cdot c + g(x) \cdot d = m$, in which $x = n$, $B_n'' = f(x)$, and $B_n''' = g(x)$.

Transposing, we have

(11) $$f_0 + n\delta_{1/2}' = y_n - B_n''(M_0'' + M_1'') - B_n'''\delta_{1/2}'''$$

or, numerically

(12) $$0.34202 + n(0.15798) = 0.38675 - B_n''(-0.02572) - B_n'''(-0.00480).$$

Applying the method, and using two calculators, one of which evaluates the left and the other the right side of the equation, we have for nearest two digits in CR

First calculator:

	PR		SMA	CR		PR
Steps 1-2	0.34202	+	(0.15798)	(0.28)	=	0.3862544.

First approximate $n = 0.28$; corresponding $B'' \doteq -0.05$, $B''' = 0.007$.

It is also noted that $B_n'''\delta_{1/2}''' = -0.00003$, when $n = 0.28$, so it is unlikely that any probable change in n from its first approximation of 0.28 would significantly affect $B_n'''\delta_{1/2}'''$. Should this not be the case, it is advisable to estimate roughly the probable value of $B_n''(M_0'' + M_1'') + B_n'''\delta_{1/2}'''$ for n equal to its first approximation, thus roughly evaluating the right-hand side of the equation. Then adjust CR of first calculator until its PR equals such value. Its CR will then show a revised n from which a closer B_n''' may be obtained. It is particularly to be desired that the coefficient B_n''' be entered in the second calculator in the step below for as close to its final value as is possible, thus avoiding lengthy final adjustment of the balancing.

Second calculator:

	PR		SMA	CR		PR
Intermediate step	0.38675	+	(0.00480)	(0.007)	=	0.3867836.

NOTE: The second term is entered in the manner to increase reading of PR as required by the signs of its factors in (12).

	PR		SMA	CR		PR
Step 3	0 3867836	+	(0.02572)	(−0.050)	=	0.3854976.

NOTE: The second term is entered in a manner to decrease reading of PR as required by the signs of its factors in (12), B'' being minus.

First calculator has PR equalized to that of second calculator, thus:

	PR		SMA	CR		PR
Step 4	0.3862544	+	(0.15798)	(0.28 to 0.27521)	=	0.3854976758.

Second approximate $n = 0.27521$, corresponding $B_n'' = -0.0499$, B_n''' unchanged.

Second calculator has CR altered to new B_n''

	PR		SMA	CR		PR
Step 5	0.3854976	+	(0.02572)	(−0.050 to −0.0499)	=	0.385500172.

First calculator has PR equalized to that of second calculator, thus:

Repeat	PR		SMA		CR		PR
Step 4	0.3854976758	+	(0.15798)		(0.27521 to 0.275226)	=	0.385500203.

Third approximate $n = 0.27523$; B_n'' and B_n''' unchanged.

As x_n is desired to only five figures (which is as great as permissible if the tabulated y's have final digits rounded; i.e., with an error of $< \pm 0.000005$) the desired x_n is 2.2752.

27. Building functions by integration from their higher differences when knowing the "lead" value of starting differences of all intermediate orders is particularly within the province of some accounting machines which have the property of transferring amounts automatically from one accumulating register to another, either additively or subtractively. The National and Underwood accounting machines are notable in this respect. However, the rotary calculator readily handles numerous problems involving second differences.

The method is best shown by an example:

x	y	1st diff.	2nd diff.
2.0	7.4079		
		0.073664	
2.1	7.4816		.006088
2.2	7.5614		.006211
2.3 etc.	7.6474 etc.		.006334 etc.

When given the lead values of y, and first difference and second difference together with the column of the latter, find the values of the second integral; i.e., the column of y's.

(1) By suitable means enter leading values in calculator as follows:

NOTE: The entry at left of SMA is leading first diff. rounded to four figures.

CR	shows	2.0
PR	"	7.4079.073664
SMA	"	.0737.006088

(2) Multiply by .1 and then change SMA to next first and second differences, reading the former from right of PR, after which

CR	shows	2.1
PR	"	7.4816.079752
SMA	"	.0798.006211

and so forth.

Obviously, the second differences should be exact if possible, and the rounding to fewer figures taken in first differences and the desired function.

28. As automatic calculators have means of adding or subtracting in both PR and CR, *in the latter by multiplying with empty SMA*, and are especially suitable for dividing subsequent to a multiplication, computing work may often be simplified to a continuous series of calculator operations by making such alterations in the procedure to adapt it best to calculator techniques.

The general expression of the best form for such computations is

$$\frac{a \pm b \cdots\cdots (\pm cd \pm ef \pm \cdots\cdots)}{\pm g} \pm h \pm i \cdots \tag{13}$$

in which the order of the entries is alphabetical and h, i, $\cdots$ are added or subtracted in CR with 1 set in SMA to show proof in PR of CR entry.

By means described in Sec. 5, even this expression may be retained in the calculator for further operations by transferring it to PR by multiplying a setup of 1 in SMA by the amount in CR (reversed to reduce it to zeros).

If there is much computing of one type, the method that it is proposed to use should be explored to see if modifications may reduce it to the above standard form, with resulting elimination of copying intermediate amounts, resetting, etc.

Because the modern calculating machine will multiply substantially as rapidly as the factors may be entered into keyboards, many long established computing techniques, such as the logarithmic, are giving way to direct computations.

A typical example is the revival of the Vieta method of approximating the roots of integral functions (polynomials). It gave way to the complete Ruffini-Horner modification principally because of the latter's use of small multipliers for obtaining the successive "reduced" equations. Because the calculating machine imposes substantially no limit to length of multiplier, the original Vieta method as modified by Birge[4] (or the Newton-Raphson) is now to be preferred for many reasons, among which is that the return to the original coefficients at each step eliminates the possibility of cumulative errors should any of the coefficients of the R-H "reduced" equations be incorrectly computed.

TABLE 1

Divisors for Use in Extracting Five-Figure Square Roots

For description, see page 60.

This table, copyrighted 1940 by Marchant Calculating Machine Company, Oakland, Calif., U.S.A., is reproduced by its express permission. The table is also available with reciprocals of the amounts shown in Cols. 1 and 2, respectively for use when the multiplication process is more suitable than that of division.

100-187			190-334		
A	**Col. 1**	**Col. 2**	**A**	**Col. 1**	**Col. 2**
100	2 000 000	6 324 555	190	2 756 810	8 717 798
102	2 019 901	6 387 488	193	2 778 489	8 786 353
104	2 039 608	6 449 806	196	2 800 000	8 854 377
106	2 059 126	6 511 528	199	2 821 347	8 921 883
108	2 078 461	6 572 671	202	2 842 534	8 988 882
110	2 097 618	6 633 250	205	2 863 564	9 055 385
112	2 116 601	6 693 280	208	2 884 441	9 121 403
114	2 135 416	6 752 777	211	2 905 168	9 186 947
116	2 154 066	6 811 755	214	2 925 748	9 252 027
118	2 172 556	6 870 226	217	2 946 184	9 316 652
120	2 190 890	6 928 203	220	2 966 479	9 380 832
122	2 209 072	6 985 700	223	2 986 637	9 444 575
124	2 227 104	7 042 727	226	3 006 659	9 507 891
126	2 244 994	7 099 296	229	3 026 549	9 570 789
128	2 262 742	7 155 418	232	3 046 309	9 633 276
130	2 280 351	7 211 103	235	3 065 942	9 695 360
132	2 297 825	7 266 361	239	3 091 925	9 777 525
134	2 315 167	7 321 202	243	3 177 691	9 859 006
136	2 332 381	7 375 636	247	3 143 247	9 939 819
138	2 349 468	7 429 670	251	3 168 596	1 001 998
140	2 366 432	7 483 315	255	3 193 744	1 009 951
142	2 383 277	7 536 577	259	3 218 695	1 017 841
144	2 400 000	7 589 466	263	3 243 455	1 025 671
146	2 416 609	7 641 989	267	3 268 027	1 033 441
148	2 433 105	7 694 154	271	3 292 416	1 041 153
150	2 449 490	7 745 969	275	3 316 625	1 048 809
152	2 465 766	7 797 435	279	3 340 659	1 056 409
154	2 481 935	7 848 567	283	3 364 521	1 063 955
156	2 497 999	7 899 367	287	3 388 215	1 071 448
158	2 513 961	7 949 843	291	3 411 744	1 078 888
160	2 529 822	8 000 000	295	3 435 112	1 086 278
163	2 553 429	8 074 652	299	3 458 323	1 093 618
166	2 576 820	8 148 620	303	3 481 379	1 100 909
169	2 600 000	8 221 922	307	3 504 283	1 108 152
172	2 622 975	8 294 577	311	3 527 038	1 115 348
175	2 645 751	8 366 600	315	3 549 648	1 122 497
178	2 668 333	8 438 009	319	3 572 115	1 129 602
181	2 690 725	8 508 819	324	3 600 000	1 138 420
184	2 712 932	8 579 044	329	3 627 671	1 147 170
187	2 734 959	8 648 699	334	3 655 133	1 155 855

339-611		
A	Col. 1	Col. 2
339	3 682 391	1 164 474
344	3 709 447	1 173 030
349	3 736 308	1 181 524
354	3 762 978	1 189 958
359	3 789 459	1 198 332
364	3 815 757	1 206 648
369	3 841 875	1 214 907
374	3 867 816	1 223 111
379	3 893 584	1 231 260
384	3 919 184	1 239 355
389	3 944 617	1 247 397
394	3 969 887	1 255 388
399	3 994 997	1 263 329
405	4 024 922	1 272 792
411	4 054 627	1 282 186
417	4 084 116	1 291 511
423	4 113 393	1 300 769
429	4 142 463	1 309 962
435	4 171 331	1 319 091
441	4 200 000	1 328 157
447	4 228 475	1 337 161
453	4 256 759	1 346 106
459	4 284 857	1 354 991
465	4 312 772	1 363 818
471	4 340 507	1 372 589
477	4 368 066	1 381 304
483	4 395 452	1 389 964
489	4 422 668	1 398 571
495	4 449 719	1 407 125
502	4 481 071	1 417 039
509	4 512 206	1 426 885
516	4 543 127	1 436 663
523	4 573 839	1 446 375
530	4 604 346	1 456 022
537	4 634 652	1 465 606
544	4 664 762	1 475 127
551	4 694 678	1 484 588
558	4 724 405	1 493 988
565	4 753 946	1 503 330
572	4 783 304	1 512 614
579	4 812 484	1 521 841
587	4 845 617	1 532 319
595	4 878 524	1 542 725
603	4 911 212	1 553 062
611	4 943 683	1 563 330

619-999		
A	Col. 1	Col. 2
619	4 975 942	1 573 531
627	5 007 994	1 583 667
635	5 039 841	1 593 738
643	5 071 489	1 603 746
651	5 102 940	1 613 691
659	5 134 199	1 623 576
667	5 165 269	1 633 401
675	5 196 152	1 643 168
683	5 226 854	1 652 876
692	5 261 179	1 663 731
701	5 295 281	1 674 515
710	5 329 165	1 685 230
719	5 362 835	1 695 877
728	5 396 295	1 706 458
737	5 429 549	1 716 974
746	5 462 600	1 727 426
755	5 495 453	1 737 815
764	5 528 110	1 748 142
773	5 560 576	1 758 408
782	5 592 853	1 768 615
791	5 624 944	1 778 764
801	5 660 389	1 789 972
811	5 695 612	1 801 111
821	5 730 620	1 812 181
831	5 765 414	1 823 184
841	5 800 000	1 834 121
851	5 834 381	1 844 993
861	5 868 560	1 855 802
871	5 902 542	1 866 548
881	5 936 329	1 877 232
891	5 969 925	1 887 856
901	6 003 332	1 898 420
912	6 039 868	1 909 974
923	6 076 183	1 921 458
934	6 112 283	1 932 874
945	6 148 170	1 944 222
956	6 183 850	1 955 505
967	6 219 325	1 966 723
978	6 254 598	1 977 878
989	6 289 674	1 988 970
999	6 321 392	1 999 000

TABLE 2

DIVISORS FOR USE IN EXTRACTING FIVE-FIGURE CUBE ROOTS

For description, see page 63.

This table, copyrighted 1940 and 1944 by Marchant Calculating Machine Company, Oakland, Calif., U.S.A., is reproduced by its express permission.

100-182				185-325			
A	Col. 1	Col. 2	Col. 3	A	Col. 1	Col. 2	Col. 3
100	300 000	139 248	646 330	185	452 101	209 847	974 023
102	303 987	141 098	654 920	188	456 976	212 109	984 524
104	307 948	142 937	663 453	191	461 824	214 360	994 970
106	311 883	144 763	671 932	194	466 648	216 599	1 005 362
108	315 794	146 579	680 357	197	471 446	218 826	1 015 700
110	319 681	148 383	688 731	200	476 220	221 042	1 025 986
112	323 544	150 176	697 054	203	480 971	223 247	1 036 220
114	327 384	151 958	705 328	206	485 698	225 441	1 046 404
116	331 202	153 730	713 554	209	490 402	227 624	1 056 539
118	334 998	155 492	721 732	212	495 084	229 797	1 066 625
120	338 773	157 244	729 864	215	499 743	231 960	1 076 664
122	342 527	158 987	737 951	218	504 381	234 113	1 086 656
124	346 260	160 720	745 995	221	508 998	236 256	1 096 603
126	349 973	162 443	753 995	224	513 594	238 389	1 106 505
128	353 667	164 158	761 953	227	518 169	240 513	1 116 362
130	357 342	165 863	769 869	230	522 725	242 627	1 126 177
132	360 997	167 560	777 745	233	527 260	244 733	1 135 948
134	364 635	169 248	785 581	236	531 777	246 829	1 145 678
136	368 254	170 928	793 379	240	537 769	249 610	1 158 587
138	371 855	172 600	801 138	244	543 727	252 376	1 171 425
140	375 439	174 264	808 860	248	549 654	255 127	1 184 193
142	379 007	175 919	816 545	252	555 548	257 863	1 196 892
144	382 557	177 567	824 194	256	561 411	260 584	1 209 524
146	386 091	179 208	831 808	260	567 244	263 291	1 222 091
148	389 609	180 841	839 387	264	573 047	265 985	1 234 593
150	393 111	182 466	846 932	268	578 821	268 665	1 247 033
152	396 598	184 084	854 444	272	584 566	271 332	1 259 410
154	400 069	185 696	861 923	276	590 284	273 985	1 271 727
156	403 526	187 300	869 369	280	595 973	276 626	1 283 985
158	406 967	188 897	876 784	284	601 636	279 254	1 296 184
160	410 394	190 488	884 168	288	607 272	281 870	1 308 327
162	413 807	192 072	891 520	292	612 881	284 474	1 320 413
164	417 206	193 650	898 843	296	618 466	287 066	1 332 444
166	420 591	195 221	906 136	300	624 025	289 647	1 344 421
168	423 962	196 786	913 400	304	629 560	292 216	1 356 345
170	427 321	198 345	920 634	308	635 070	294 773	1 368 217
173	432 333	200 671	931 434	312	640 557	297 320	1 380 038
176	437 317	202 985	942 171	316	646 020	299 856	1 391 808
179	442 273	205 285	952 847	320	651 460	302 381	1 403 529
182	447 200	207 572	963 464	325	658 229	305 523	1 418 111

330-614				622-1000			
A	Col. 1	Col. 2	Col. 3	A	Col. 1	Col. 2	Col. 3
330	664 963	308 648	1 432 618	622	101 465	470 957	2 185 990
335	671 663	311 758	1 447 053	631	102 441	475 489	2 207 026
340	678 329	314 853	1 461 416	640	103 413	480 000	2 227 963
345	684 963	317 932	1 475 708	649	104 380	484 490	2 248 801
350	691 565	320 996	1 489 933	658	105 343	488 958	2 269 544
355	698 136	324 046	1 504 088	667	106 301	493 407	2 290 192
360	704 676	327 082	1 518 179	676	107 255	497 835	2 310 747
365	711 186	330 103	1 532 204	685	108 205	502 244	2 331 211
370	717 666	333 111	1 546 165	694	109 151	506 634	2 351 586
375	724 117	336 105	1 560 063	703	110 093	511 005	2 371 873
380	730 539	339 086	1 573 899	712	111 030	515 357	2 392 074
386	738 209	342 646	1 590 424	721	111 964	519 690	2 412 190
392	745 839	346 188	1 606 862	730	112 894	524 006	2 432 220
398	753 431	349 712	1 623 217	740	113 922	528 781	2 454 383
404	760 984	353 217	1 639 490	750	114 946	533 534	2 476 445
410	768 500	356 706	1 655 683	760	115 966	538 266	2 498 410
416	775 979	360 178	1 671 797	770	116 981	542 977	2 520 278
422	783 423	363 633	1 687 833	780	117 992	547 668	2 542 051
428	790 831	367 071	1 703 794	790	118 998	552 339	2 563 732
434	798 205	370 494	1 719 680	800	120 000	556 991	2 585 322
440	805 545	373 901	1 735 494	811	121 097	562 085	2 608 966
446	812 851	377 292	1 751 235	822	122 190	567 156	2 632 505
452	820 125	380 668	1 766 906	833	123 278	572 204	2 655 938
458	827 367	384 030	1 782 508	844	124 361	577 231	2 679 268
465	835 776	387 933	1 800 625	855	125 439	582 235	2 702 498
472	844 143	391 816	1 818 651	866	126 512	587 219	2 725 628
479	852 468	395 681	1 836 587	877	127 581	592 181	2 748 660
486	860 753	399 526	1 854 437	888	128 646	597 122	2 771 596
493	868 999	403 354	1 872 201	900	129 802	602 490	2 796 509
500	877 205	407 163	1 889 882	912	130 954	607 833	2 821 312
507	885 374	410 954	1 907 480	924	132 100	613 154	2 846 007
514	893 504	414 728	1 924 997	936	133 241	618 451	2 870 594
521	901 598	418 485	1 942 435	948	134 378	623 725	2 895 077
528	909 656	422 225	1 959 795	961	135 603	629 415	2 921 484
535	917 678	425 949	1 977 078	974	136 823	635 078	2 947 772
542	925 666	429 656	1 994 286	987	138 038	640 717	2 973 943
550	934 752	433 873	2 013 862	1000	139 248	646 330	3 000 000
558	943 794	438 071	2 033 343				
566	952 794	442 248	2 052 732				
574	961 751	446 405	2 072 029				
582	970 666	450 543	2 091 237				
590	979 541	454 663	2 110 357				
598	988 376	458 763	2 129 391				
606	997 171	462 846	2 148 340				
614	100 593	466 910	2 167 206				

NOTES

1. Written by Tracy W. Simpson, Sc. B., E.E., Mem. Amer. Math. Soc. To reflect current methods with American-made calculators, this section has been completely rewritten, including an extensive outline of calculator applications. These have no counterpart in the German edition.

2. For a discussion of equations of this type, see Art. 10.

3. Cf. *Direct Interpolation and Subtabulation*, published as MM-189 by Marchant Calculating Machine Company, also loc. cit. Note 7 (below).

4. Cf. *The Birge-Vieta Method for Real Roots of Rational Integral Functions*, published as MM-225 by Marchant Calculating Machine Company.

5. See Barlow's *Tables*, Introduction by L. J. Comrie.

6. Pamphlet MM-222, Marchant Calculating Machine Company.

7. *Interpolation and Allied Tables*, by L. J. Comrie, H. M. Stationery Office, London W. C. 2; and MM-220 and MM-221, pamphlets published by Marchant Calculating Machine Company.

CHAPTER TWO

INTERPOLATION

7. The Rational Integral Function.

1. In this chapter we shall treat with the following *problem*: A number of related values of two or more variables are given; a function is sought which reproduces this relation with sufficient accuracy. If we know nothing more than this about the theoretical relation of the variables, it is natural for us to choose as a representation of this dependence a function which, mathematically speaking, is as easy to handle as possible. Such a function is the rational integral function. If, for a function of one variable, we consider the given values as abscissas and ordinates in a rectangular coordinate system, then our task is to draw a polynomial curve of some order through these points. In this case the given values are assumed to be exact. If only two points are known, the interpolation curve is a straight line. This case has already been treated in Art. 4. To become familiar with the rational integral function, we will first assume that the desired rational function has been found and that several values of this function are to be calculated or constructed.

2. For calculation purposes, it is customary to write the function

$$y_n(x) = a_n x^n + a_{n-1} x^{n-1} \cdots + a_1 x + a_0$$

in a way similar to decimals:

$$a_n\,,\ a_{n-1}\,,\ a_{n-2} \cdots a_2\,,\ a_1\,,\ a_0\,,$$

i.e., by omission of the powers of x. The index of the coefficient denotes the exponent of the particular power of x. If the coefficients are given numerical values, we must not forget to insert zeros in the appropriate places of this scheme. For example, instead of $y = x^5 - 3x^2 + 5$ we write

$$1, 0, 0, -3, 0, +5.$$

In calculating the value of such a function for $x = x_0$, we do not calculate the individual powers, multiply by the respective coefficients and add. Instead, we calculate the following expressions successively:

$$\begin{aligned}
a'_{n-1} &= a_n x_0 + a_{n-1}\,, \\
a'_{n-2} &= a'_{n-1}x_0 + a_{n-2} = a_n x_0^2 + a_{n-1}x_0 + a_{n-2}\,, \\
&\cdots\cdots\cdots\cdots\cdots\cdots \\
a'_0 &= a'_1 x_0 + a_0 \qquad = a_n x_0^n + a_{n-1}x_0^{n-1} \cdots a_1 x_0 + a_0 \\
&= g(x_0).
\end{aligned} \tag{1}$$

These calculations are best carried out in the form of a scheme given by Horner[1]:

$$\begin{array}{cccccccc}
a_n & a_{n-1} & a_{n-2} & a_{n-3} & \cdots & a_2 & a_1 & a_0 \\
 & x_0 a_n & x_0 a'_{n-1} & x_0 a'_{n-2} & \cdots & x_0 a'_3 & x_0 a'_2 & x_0 a'_1 \\
\hline
 & a'_{n-1} & a'_{n-2} & a'_{n-3} & \cdots & a'_2 & a'_1 & \underline{a'_0 = g(x_0).}
\end{array} \tag{2}$$

The advantage of this type of calculation is that in the same sequence we always carry out multiplication by x_0 and addition. If the accuracy of the slide rule suffices, only a single setting of the slide is necessary. If a calculating machine is available, only the single multiplication factor x_0 need be set on the machine.

Exactly the same calculations occur if $g(x)$ is divided by $(x - x_0)$, as is seen in the boxed expressions.

$$\begin{array}{l}
\left(\boxed{a_n}\, x^n + \boxed{a_{n-1}}\, x^{n-1} + a_{n-2}\, x^{n-2} + \cdots + a_2 x^2 + a_1 x + a_0\right) : (x - x_0) = a_n x^{n-1} \\
\underline{a_n\ \ x^n - \boxed{x_0 a_n}\, x^{n-1}} \qquad\qquad\qquad\qquad + a'_{n-1}x^{n-2} + a'_{n-2}x^{n-3} + \cdots + a'_2 x + a'_1 \\
\qquad\quad \boxed{a'_{n-1}}\, x^{n-1} + \boxed{a_{n-2}}\, x^{n-2} \\
\qquad\quad \underline{a'_{n-1}\ \ x^{n-1} - \boxed{a'_{n-1}x_0}\, x^{n-2}} \\
\qquad\qquad\qquad\qquad \boxed{a'_{n-2}}\, x^{n-2} \cdots \\
\qquad\qquad\qquad\qquad \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\
\qquad\qquad\qquad\qquad\qquad\qquad\qquad \boxed{a'_1}\, x + \boxed{a_0} \\
\qquad\qquad\qquad\qquad\qquad\qquad\qquad \underline{a'_1\ \ x - \boxed{x_0 a'_1}} \\
\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad \boxed{a'_0}
\end{array} \tag{2a}$$

From this it follows that we can write

$$(3)\quad g(x) = a_0' + (x - x_0)[a_1' + a_2'x + a_3'x^2 + \cdots + a_{n-1}'x^{n-2} + a_nx^{n-1}].$$

Therefore, $a_0' = g(x_0)$. It is customary to write the symbol $[xx_0]$ for the bracketed equation [v. 8(11)]. We can now divide $[xx_0]$ by another value $(x - x_1)$.

$$(4)\quad \begin{array}{cccccc|l} a_n & a_{n-1}' & a_{n-2}' & \cdots & a_2' & a_1' & \\ & x_1a_n & x_1a_{n-1}'' & \cdots & x_1a_3'' & x_1a_2'' & \\ \hline & a_{n-1}'' & a_{n-2}'' & & a_2'' & & \underline{a_1'' = g_1(x_1).} \end{array}$$

This becomes

$$(5)\quad \begin{aligned} g(x) &= a_0' + a_1''(x - x_0) \\ &\quad + (x - x_0)(x - x_1)[a_2'' + a_3''\cdot x + \cdots + a_{n-1}''x^{n-3} + a_nx^{n-2}], \end{aligned}$$

in which $a_1'' = [x_1x_0]$ and the function in the brackets is denoted by $[xx_1x_0]$.

While $[xx_0]$ was of the $(n - 1)$st degree, $[xx_1x_0]$ is of the $(n - 2)$nd degree. If we continue in this way, we obtain a product expansion of the rational integral function

$$(6)\quad \begin{aligned} g(x) &= a_0' + a_1''(x - x_0) + a_2'''(x - x_0)(x - x_1) \\ &\quad + a_3''''(x - x_0)(x - x_1)(x - x_2) + \cdots \\ &\quad + a_n(x - x_0)(x - x_1) \cdots (x - x_{n-1}), \end{aligned}$$

where $[xx_{n-1} \cdots x_1x_0] = a_n$ is a constant.

If one should calculate the value of such a product development for any argument $\bar{x}$, it is advisable to use a scheme corresponding to Horner's scheme, in which only the factor changes from column to column:

$$(7)\quad \begin{array}{cccccc} a_n \quad a_{n-1}^{(n)} & a_{n-2}^{(n-1)} & \cdots & a_2''' & a_1'' & a_0' \\ + a_n(\bar{x} - x_{n-1}) & + b_{n-1}(\bar{x} - x_{n-2}) & \cdots & + b_3(\bar{x} - x_2) & b_2(\bar{x} - x_1) & b_1(\bar{x} - x_0) \\ \hline b_{n-1} & b_{n-2} & & b_2 & b_1 & b_0 = g(\bar{x}) \end{array}$$

This arrangement needs no further explanation.

3. *Example*: The function $y = x^3 - 3x^2 + 1$ is to be developed in an expression in which the factors $(x - 0.5)$, $(x + 0.5)$ and $(x - 3)$ appear successively. First we form Horner's scheme:

$$
\begin{array}{r|rrr|l}
 & 1 & -3 & 0 & 1 \\
+0.5: & & +0.5 & -1.25 & -0.625 \\
\hline
 & 1 & -2.5 & -1.25 & \boxed{+0.375 = a_0'} \\
-0.5: & & -0.5 & +1.5 & \\
\hline
 & 1 & -3 & \boxed{+0.25 = a_1''} & \\
+3 & & +3 & & \\
\hline
 & 1 & \boxed{0 = a_2'''.} & &
\end{array}
$$

Then we obtain

$$y = 0.375 + 0.25(x - 0.5) + (x - 0.5)(x + 0.5)(x - 3).$$

If the value of this expression is calculated for $\bar{x} = 2$, then the scheme

$$
\begin{array}{ccc}
\bar{x} - 3 & \bar{x} + 0.5 & \bar{x} - 0.5 \\
\hline
0 & 0.25 & 0.375 \\
-1 & -2.5 & -3.375 \\
\hline
-1 & -2.25 & \boxed{-3.}
\end{array}
$$

is obtained.

4. If we set $x_0 = x_1 = \cdots = x_{n-1}$, then we have a power series instead of a product expansion:

$$(8) \quad g_n(x) = a_0' + a_1''(x - x_0) + a_2'''(x - x_0)^2 + \cdots + a_n(x - x_0)^n.$$

This is of less importance for interpolation than for the calculation of the roots of the rational integral function (cf. Art. 19).

Example: The general equation of the third degree

$$x^3 + a_2x^2 + a_1x + a_0 = 0$$

is to be transformed to the reduced form, in which the coefficient of the second power of the unknown is zero. We introduce the new variable $z = x + a_2/3$; i.e., the function is developed in powers of $x + a_2/3$. For example, in the equation

$$x^3 - 5.7x^2 + 1.32x + 6.21 = 0,$$

$a_2/3 = -1.9$, and the following development results:

$$\begin{array}{cccc}
1 & -5.7 & +1.32 & +\ 6.21 \\
 & +1.9 & -7.22 & -11.21 \\
\hline
1 & -3.8 & -5.90 & \multicolumn{1}{|l}{-\ 5.00 = a_0'} \\
\cline{4-4}
 & +1.9 & -3.61 & \\
\cline{1-3}
1 & -1.9 & \multicolumn{1}{|l}{-9.51 = a_1''} & \\
\cline{3-3}
 & +1.9 & & \\
\cline{1-2}
1 & \multicolumn{1}{|l}{0 = a_2'''.} & & \\
\cline{2-2}
\end{array}$$

Then the equation will be

$$(x - 1.9)^3 - 9.51(x - 1.9) - 5 = 0.$$

5. The *Horner scheme* can be put *into graphical form*, and can be used both for the construction of the curve representing the function $y = g(x)$, and for the construction of the corresponding scale. The point-by-point construction of the curve is carried out by Segner[2] in the following way.

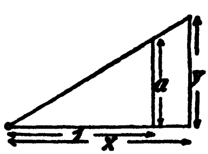


FIG. 20

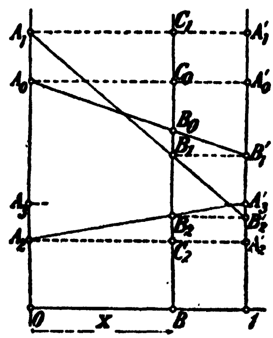


FIG. 21

From Fig. 20, it follows, by similar triangles, that $y = ax$. Segner used this fact in his construction. For example, if a function of third degree

(9) $$y = a_3x^3 + a_2x^2 + a_1x + a_0$$

is to be constructed, we lay off the lengths

$$\text{(9a)}\qquad \begin{aligned} 0A_0 &= 1A_0' = a_0 \ , \ A_0A_1 = A_0'A_1' = a_1 \ , \\ A_1A_2 &= A_1'A_2' = a_2 \ , \ A_2A_3 = A_2'A_3' = a_3 \ , \end{aligned}$$

on the y axis. These are marked off in the positive or negative direction according to the sign. (In Fig. 21, which corresponds to the example in Section 4, $A_1A_2 = a_2$ is negative.) In the drawing,

$$C_2B_2 = a_3 \cdot x; \ C_1B_2 = a_3x + a_2 = a_2' = A_1'B_2'$$

$$\text{(10)}\qquad C_1B_1 = (a_3x + a_2)x; \ C_0B_1 = a_3x^2 + a_2x + a_1 = a_1' = A_0'B_1'$$

$$C_0B_0 = (a_3x^2 + a_2x + a_1)x; \ BB_0 = a_3x^3 + a_2x^2 + a_1x + a_0 = a_0' \ .$$

The value B_0 is then the point of the curve belonging to the abscissa $OB = 2/3$. This construction has been further developed by Massau (v. 9.3).

6. The Horner scheme has been put into graphical form in another way by Lill[3] for the construction of the function in scale form. He makes use of the fact that in the arrangement of the various lengths shown in Fig.

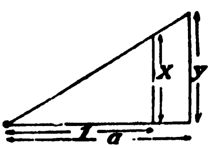


FIG. 22

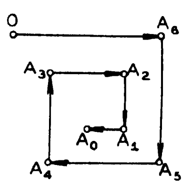


FIG. 23

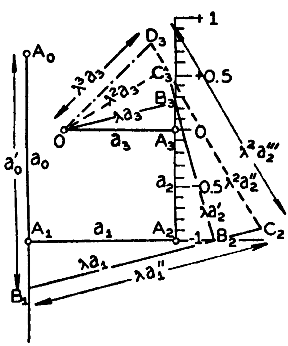


FIG. 24

22, $y = ax$. Lill then plots successively the lengths representing the coefficients, beginning with the coefficient of the highest power. These are plotted so that each is perpendicular to the previous one, those with positive coefficients turning clockwise, and those with negative coefficients counter-clockwise. (The scheme is shown in Fig. 23 for positive terms only. In Fig. 25, which corresponds to the preceding example, a_2 is negative.) If the coefficient of a power is zero, we must note that, in the transition to the next coefficient, a rotation of 180° must be performed. In

Fig. 24, $OA_3A_2A_1A_0$ is the polygon representing the function

$$y = a_3x^3 + a_2x^2 + a_1x + a_0 .$$

In this framework, another polygon is now drawn, the segments of which are successively perpendicular, and which have the slope $\tan \varphi = x_0$ with respect to the corresponding side of the original polygon. The value x_0 is the value for which the function is to be calculated. The broken line $OB_3B_2B_1$ is such a polygon for $x_0 = 1/4$. In this case,

$$(11)\qquad \begin{aligned} A_3B_3 &= a_3x_0 \ ; \ A_2B_3 = a_3x_0 + a_2 = a_2' \ , \\ A_2B_2 &= (a_3x_0 + a_2)x_0 \ ; \ A_1B_2 = a_3x_0^2 + a_2x_0 + a_1 = a_1' \ , \\ A_1B_1 &= (a_3x_0^2 + a_2x_0 + a_1)x_0 \ ; \\ A_0B_1 &= a_3x_0^3 + a_3x_0^2 + a_2x_0 + a_0 = a_0' \ . \end{aligned}$$

Therefore A_0B_1 is the desired value of the function. The sign of the length $A_{n-1}B_n$ is to be taken as positive if A_nB_n has the direction $A_{n-1}A_n$ for positive coefficients.

7. The advantage of this representation is that it permits a graphical determination of the coefficients of the expansion of a rational integral function in products $(x - x_0)$, $(x - x_0)(x - x_1) \cdots$ or in powers of $(x - x_0)$. We consider first the power series expansion. By similar triangles,

$$\Delta OA_3B_3 \sim \Delta B_3A_2B_2 \sim \Delta B_2A_1B_1 .$$

If we let $\lambda = (1 + x_0^2)^{\frac{1}{2}}$,

$$OB_3 = \lambda a_3 \ ; \qquad B_3B_2 = \lambda a_2' \ ; \qquad B_2B_1 = \lambda a_1' \ ;$$

i.e., if the unit length is taken λ times as large, then $OB_3B_2B_1$ is the polygon of the function in brackets in the expression

$$(12)\qquad y = a_0' + (x - x_0)[a_1' + a_2'x + a_3x^2].$$

If we draw the rectangular polygon in this framework for $x = x_0$, starting out again from the length OB_3, we obtain the points OC_3C_2, and

$$(13)\qquad \begin{aligned} B_2C_3 &= \lambda a_3x_0 \ ; & B_2C_3 &= \lambda(a_3x_0 + a_2') = \lambda a_2'' ; \\ B_2C_2 &= \lambda(a_3x_0 + a_2')x_0 \ ; & B_1C_2 &= \lambda(a_3x_0^2 + a_2'x_0 + a_1') = \lambda a_1'' . \end{aligned}$$

By similar triangles,

$$\Delta OA_3B_3 \sim \Delta OB_3C_3 \sim \Delta C_3B_2C_2 ,$$

and further,

$$OC_3 = \lambda^2 a_3, \qquad C_3C_2 = \lambda^2 a_2'',$$

that is, if

$$(14) \qquad y = a_0' + a_1''(x - x_0) + (x - x_0)^2[a_2'' + a_3 x],$$

then the polygon OC_3C_2 is the sequence of lines of the polygon represented in the brackets, except that the unit length here is λ^2, etc.

The polygon $OD_3C_2B_1A_0$ gives the coefficients of the expansion of y in powers of $x - x_0$. Of course, each length must be measured with a different scale:

$$OD_3 = \lambda^3 a_3; \quad D_3C_2 = \lambda^2 a_2'''; \quad C_2B_1 = \lambda a_1''; \quad B_1A_0 = a_0'.$$

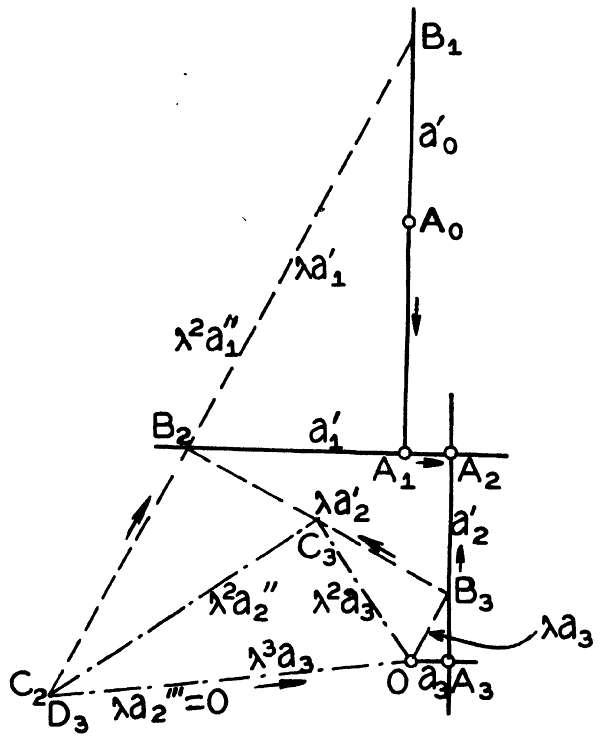


FIG. 25

In Fig. 25, which represents the example of Section 4, $D_3C_2 = 0$, corresponding to the expansion

$$y = -5 - 9.51(x - 1.9) + (x - 1.9)^3.$$

8. To determine the coefficients of the product expansion, we would

draw the polygon for $x = x_0$ about the first polygon, the polygon for $x = x_1$ in the framework thus obtained, etc.

If it is desired to construct the scale for such a product expansion, the coefficients must be joined to a framework exactly as in the power series. In this case, however, we do not construct a rectangular polygon. Instead we construct a polygon the sides of which must have the direction, with respect to the corresponding side of the framework, given by

$$\operatorname{tg} \varphi_n = (x - x_n), \qquad \operatorname{tg} \varphi_{n-1} = x - x_{n-1}\text{, etc.}$$

9. The coefficients of the product expansion can also be found from those of the power series by means of *Segner's construction*, 7 (5). However, this will not be considered here. The construction of the curve representing the function, as given by *Massau* in a generalization of Segner's construction, is more convenient than that of the scale according to the generalized method of Lill.

As in the Segner method, the coefficients of the product expansion are plotted successively on the y axis, the positive coefficients in the positive direction of the axis and vice versa. This is shown schematically in Fig. 26 for a function of third degree. We draw lines parallel to the y axis through the abscissas x_0, x_1, x_2 and $x_0 + 1$, $x_1 + 1$, $x_2 + 1$ as well as through the abscissa x for which the value of the function is to be constructed. The endpoints A_2A_3 of a_3 are projected on the y-parallels through x_2 and $x_2 + 1$ to A_2' and A_3'', respectively. The line connecting these points intersects the parallel through x in B_2. Then

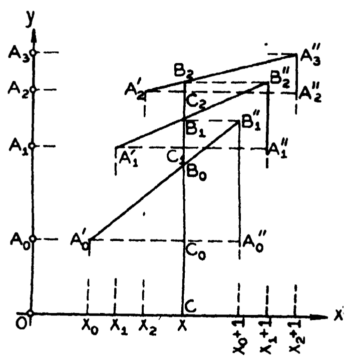


FIG. 26

$$C_2B_2 = a_3(x - x_2);$$

$$C_1B_2 = A_1''B_2'' = a_3(x - x_2) + a_2\,.$$

If we connect B_2'' on the parallel through $x_1 + 1$, with the projection A_1' of A_1 on the parallel through x_1, we have

$$C_1B_1 = [a_3(x - x_2) + a_2](x - x_1)$$

and

$$C_0B_1 = A_0''B_1'' = a_3(x - x_2)(x - x_1) + a_2(x - x_1) + a_1\,.$$

The procedure is continued in this way until we have, finally,

$$CB_0 = a_3(x - x_2)(x - x_1)(x - x_0) + a_2(x - x_1)(x - x_0) + a_1(x - x_0) + a_0 .$$

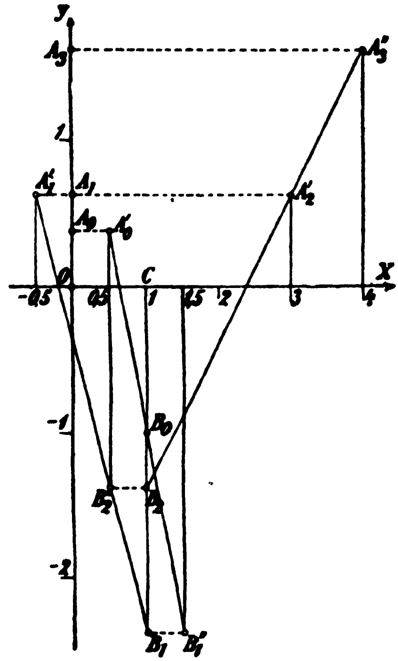


FIG. 27

Figure 27 shows the construction of the ordinate for the abscissa value $x = 1$ in the preceding example:

$$y = 0.375 + 0.25(x - 0.5) + (x - 0.5)(x + 0.5)(x - 3).$$

The result is then $CB_0 = -1$ (cf. Art. 19).

NOTES

1. Horner, *Phil. Transactions* (1819) I, p. 308
2. Segner, *Petrop. Novi Comment.* VII (1761).
3. Lill, *Nouv. Ann. de math.*, 2nd series, VI and VII (1867-68).

8. The General Interpolation Formula.

1. To obtain the product expansion of an arbitrary function, corresponding to the development given in Art 7 for the rational integral function, and simultaneously to obtain the equation of the curve which passes through a number of given points, we start out with the *so-called divided differences.*

If the values of a function $f(x)$ are given for a number of arguments $x_0, x_1, \cdots, x_n$, then the expression

$$f\frac{f(x_m) - f(x_r)}{x_m - x_r} = [x_m x_r] \tag{1}$$

is called the *divided difference,* or difference coefficient, *of first order.* Geometrically, this is the slope of the secant through the points with the abscissa x_m and x_r, if x_m and x_r are real values and $f(x)$ is a real function. From the equation

$$[x_m x_r] = \frac{f(x_m)}{x_m - x_r} + \frac{f(x_r)}{x_r - x_m} = [x_r x_m] \tag{2}$$

it may be seen that the divided difference of first order is a symmetric function of its argument, so that the order of the arguments is immaterial.

From the divided difference of the first order, the *divided difference of the second order* can be formed by a corresponding process of calculation. The expression is

$$\begin{aligned} [x_l x_m x_r] &= \frac{[x_l x_m] - [x_m x_r]}{x_l - x_r} = \frac{f(x_l) - f(x_m)}{(x_l - x_m)(x_l - x_r)} - \frac{f(x_m) - f(x_r)}{(x_m - x_r)(x_l - x_r)} \\ &= \frac{f(x_l)}{(x_l - x_m)(x_l - x_r)} + \frac{f(x_m)}{(x_m - x_l)(x_m - x_r)} + \frac{f(x_r)}{(x_r - x_l)(x_r - x_m)}. \end{aligned} \tag{3}$$

Hence the difference of two divided differences of first order, which have one argument value in common, is divided by the difference of the other argument values. It can also be seen from the above equation [illegible] the order of the arguments is arbitrary in the divided difference of second order, so that

$$[x_m x_r x_l] = [x_m x_l x_r] = [x_l x_m x_r] = [x_l x_r x_m] = [x_r x_l x_m] = [x_r x_m x_l]. \tag{4}$$

Divided differences of any higher order can be calculated in the same way. For example, the divided difference of third order is

$$[x_i x_l x_m x_r] = \frac{[x_i x_l x_m] - [x_l x_m x_r]}{x_i - x_r}. \tag{5}$$

Therefore, to form the divided difference of nth order, we first form the difference of two divided differences of the $(n - 1)$st order:

$$[x_0 x_1 \cdots x_{j-1} x_j x_{j+1} \cdots x_{n-1}] \text{ and } [x_0 x_1 \cdots x_{j-1} x_k x_{j+1} \cdots x_{n-1}].$$

Then only two argument values differ from each other. The expression above is then divided by the difference of the two argument values x_j and x_k. This is the divided difference of the nth order.

2. Now it can be shown by induction that the *differences are always symmetric functions of their argument.* It is assumed that this theorem is proved for the nth difference. Let $[x_0x_1 \cdots x_n]$ be a symmetric function of its argument, and let it be put in the form

$$[x_0x_1 \cdots x_n] = \sum_{r=0}^{n} \frac{f(x_r)}{(x_r - x_0)(x_r - x_1) \cdots (x_r - x_{r-1})(x_r - x_{r+1}) \cdots (x_r - x_n)}. \tag{6}$$

Then the $(n+1)$st difference is

$$[x_0x_1 \cdots x_{n+1}] = \frac{[x_0x_1 \cdots x_n] - [x_1x_2 \cdots x_{n+1}]}{x_0 - x_{n+1}}$$

$$\begin{aligned} &= \frac{f(x_0)}{(x_0 - x_1) \cdots (x_0 - x_n)(x_0 - x_{n+1})} \\ &\quad + \sum_{r=1}^{n} \left[\frac{f(x_r)}{(x_0 - x_{n+1})(x_r - x_0) \cdots (x_r - x_{r-1})(x_r - x_{r+1}) \cdots (x_r - x_n)} \right. \\ &\quad \left. - \frac{f(x_r)}{(x_0 - x_{n+1})(x_r - x_1) \cdots (x_r - x_{r-1})(x_r - x_{r+1}) \cdots (x_r - x_{n+1})} \right] \\ &\quad + \frac{f(x_{n+1})}{(x_{n+1} - x_0)(x_{n+1} - x_1) \cdots (x_{n+1} - x_n)}. \end{aligned} \tag{7}$$

The sum in (7) can be transformed. In place of the general term, we can write

$$\frac{f(x_r)}{(x_0 - x_{n+1})(x_r - x_1) \cdots (x_r - x_{r-1})(x_r - x_{r+1}) \cdots (x_r - x_n)} \times \left(\frac{1}{x_r - x_0} - \frac{1}{x_r - x_{n+1}} \right). \tag{8}$$

The last expression in parentheses can be written

$$\frac{1}{x_r - x_0} - \frac{1}{x_r - x_{n+1}} = \frac{x_0 - x_{n+1}}{(x_r - x_0)(x_r - x_{n+1})}. \tag{9}$$

The numerator of this fraction cancels the first factor in the denominator, so that equation (7) takes the form

$$[x_0x_1 \cdots x_{n+1}] = \sum_{r=0}^{n+1} \frac{f(x_r)}{(x_r - x_0) \cdots (x_r - x_{r-1})(x_r - x_{r+1}) \cdots (x_r - x_{n+1})}. \tag{10}$$

Then the following has been proved: If the divided difference of nth order

is written in the given form, which is independent of the order of the argument, then this is also true of the divided difference of the $(n + 1)$st order. In the previous section we showed that the divided differences of the first and second order can be put in this form. Therefore, the divided difference of any arbitrary order is a symmetric function of its argument.

It can be seen, without further discussion, that the following theorems hold for the divided differences:

The divided difference of the sum of several functions is equal to the sum of the divided differences of the several functions.

If a function is multiplied by a constant factor, then the divided differences are also multiplied by this factor.

3. We now write the equations for the divided differences in a somewhat different form:

$$
\begin{aligned}
(x - x_0)[xx_0] &= f(x) - f(x_0) \\
(x - x_1)[xx_0x_1] &= [xx_0] - [x_0x_1] \\
(x - x_2)[xx_0x_1x_2] &= [xx_0x_1] - [x_0x_1x_2] \\
&\cdots\cdots\cdots\cdots \\
(x - x_n)[xx_0 \cdots x_n] &= [xx_0 \cdots x_{n-1}] - [x_0x_1 \cdots x_n].
\end{aligned}
\tag{11}
$$

If the value for $[xx_0]$, calculated from the second equation, is substitu in the first equation, we obtain

$$
f(x) = f(x_0) + (x - x_0)[x_0x_1] + (x - x_0)(x - x_1)[xx_0x_1]. \tag{12}
$$

If the value calculated from the third equation of (11) is substitute $[xx_0x_1]$, then

$$
\begin{aligned}
f(x) = f(x_0) &+ (x - x_0)[x_0x_1] + (x - x_0)(x - x_1)[x_0x_1x_2] \\
&+ (x - x_0)(x - x_1)(x - x_2)[xx_0x_1x_2]
\end{aligned}
\tag{13}
$$

and so forth. Finally we obtain

$$
\begin{aligned}
f(x) = f(x_0) &+ (x - x_0)[x_0x_1] + (x - x_0)(x - x_1)[x_0x_1x_2] \\
&+ (x - x_0)(x - x_1)(x - x_2)[x_0x_1x_2x_3] \\
&+ \cdots + (x - x_0) \cdots (x - x_{n-1})[x_0x_1 \cdots x_n] + R_{n+1}
\end{aligned}
\tag{14}
$$

where the remainder is

$$R_{n+1} = (x - x_0)(x - x_1) \cdots (x - x_n)[xx_0x_1 \cdots x_n]. \tag{15}$$

This is the general Newton interpolation formula.[1] At first, this is only an identity. But with it we have obtained the expansion in products of an arbitrary function. The difficulty lies in the correct estimation of the remainder. The magnitude of this term determines the value of the approximation, when a finite number of terms are used for calculation. If the expansion is broken off at any term, an approximation function is obtained which has a series of discrete values in common with the function to be approximated. For example, if the series is broken off with the term $(x - x_0)(x - x_1)(x - x_2)\,[x_0x_1x_2x_3]$, then all the subsequent terms contain the factors $x - x_0$, $x - x_1$, $x - x_2$ and $x - x_3$, and for the values $x - x_0$, x_1, x_2, x_3 these terms vanish; i.e., the approximating curve has the four points with these abscissas in common with the curve $y = f(x)$.

For the rational integral function, the above expansion is identical with that obtained according to the Horner scheme.

The divided difference of nth order of a rational integral function of nth order is a constant. Each divided difference of higher order is then zero.

In other cases, the divided differences of higher order become vanishingly small, so that by neglecting the remainder, a satisfactory approximation to the path of the function is obtained in an interval $a \cdots b$, in which lie the values $x_0 \cdots x_n$.

4. *Example*: The function $y = \sin x$ is to be approximated by a curve which passes through the points with the abscissas $x = 0°$, 30°, 45°, 60°, 90°. The approximating function then has the equation

$$\begin{aligned} y = 0.016667x &- 0.000063546x(x - 30) \\ &+ 0.000000726x(x - 30)(x - 45) \\ &+ 0.0000000027x(x - 30)(x - 45)(x - 60), \end{aligned}$$

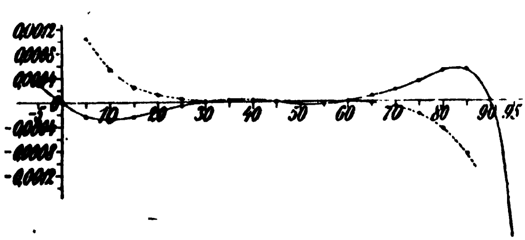


FIG. 28

where x is to be measured in degrees. The deviations from the sine

curve calculated in this way, are shown by the curve drawn in Fig. 28. The deviation between 0° and 90° is never larger than 6 in the fourth decimal place. If sin x is expanded about $x_0 = \pi/4$ in a Taylor series, we get, for $\xi = x - \pi/4$,

$$y = \frac{1}{2} 2^{\frac{1}{2}}\left(1 + \xi - \frac{\xi^2}{2} - \frac{\xi^3}{6} + \frac{\xi^4}{24} \cdots\right)$$

if we go to terms of fourth order, as was done in the product expansion. Here ξ is measured in radians. The deviations from the true value are given by the broken line in Fig. 28. Between 30° and 60°, the fifth decimal place is accurate. But the deviation becomes very large at either end of the 0°-90° interval. Thus, while the product development is accurate to three places throughout the entire interval, the Taylor series is very accurate in the vicinity of 45°, but at 5° and 85°, it is incorrect in the third decimal place, and becomes even more incorrect beyond those values.

5. Newton's general interpolation formula, which has been derived above, is only another way of writing the well-known Lagrange formula.[2] That is, if we set

$$\varphi(x) = (x - x_0)(x - x_1) \cdots (x - x_n), \tag{16}$$

then the derivative of this function consists of $n + 1$ terms of a sum, each of which is a product of n factors. The rth term of the sum is

$$(x - x_0) \cdots (x - x_{r-1})(x - x_{r+1}) \cdots (x - x_n). \tag{17}$$

Hence the factor $(x - x_r)$ is lacking, a factor which appears in all the other terms. Consequently, the derivative has the value

$$\varphi'(x_r) = (x_r - x_0) \cdots (x_r - x_{r-1})(x_r - x_{r+1}) \cdots (x_r - x_n) \tag{18}$$

for $x = x_r$. If we observe further that

$$\begin{aligned}[xx_0 \cdots x_n] = {} & \frac{f(x)}{(x-x_0) \cdots (x-x_n)} + \frac{f(x_0)}{(x_0-x)(x_0-x_1) \cdots (x_0-x_n)} + \cdots \\ & + \frac{f(x_r)}{(x_r-x)(x_r-x_0) \cdots (x_r-x_{r-1})(x_r-x_{r+1}) \cdots (x_r-x_n)} + \cdots \\ & + \frac{f(x_n)}{(x_n-x)(x_n-x_0) \cdots (x_n-x_{n-1})}\end{aligned} \tag{19}$$

as was shown in section 3, then, by application of the expressions for $\varphi(x)$ and $\varphi'(x)$, we can write

$$[xx_0 \cdots x_n] = \frac{f(x)}{\varphi(x)} + \frac{f(x_0)}{(x_0 - x)\varphi'(x_0)} + \cdots + \frac{f(x_r)}{(x_r - x)\varphi'(x_r)} + \cdots$$
$$+ \frac{f(x_n)}{(x_n - x)\varphi'(x_n)}.$$

If we solve this equation for $f(x)$, we obtain Lagrange's interpolation formula

$$f(x) = \sum_{r=0}^{n} \frac{f(x_r)\varphi(x)}{\varphi'(x_r)(x - x_r)} + [xx_0 \cdots x_n]\varphi(x), \tag{21}$$

which has a remainder identical with that of Newton's formula.

6. The observations made so far are also valid for complex values of the arguments $x_0, \cdots, x_n$, and also for complex functions. We now wish to assume that $f(x)$ *is a real function of the real variable* x, i.e., that x_0, $\cdots, x_n$ are real values of this variable. Also we assume that $f(x)$ possesses finite continuous derivatives up to the nth in the interval (a, b) being considered in the interpolation. The $n + 1$ values of the argument, $_0, \cdots, x_n$, lie in this interval. We call the approximating function, erived in section 4, (without the remainder term) $N(x)$, and form a ew function

2) $$F(x) = f(x) - N(x).$$

is function has at least $n + 1$ roots in the interval (a, b), since $f(x_0) =$ $x_0), \cdots, f(x_n) = N(x_0)$. Then, by Rolle's theorem, $F'(x)$ must have n ts in this interval, $F''(x)$ must have $n - 1$ roots, etc. $F^{(n)}(x)$ has at st one root. Then,

) $$f^{(n)}(\xi) = N^{(n)}(\xi).$$

t, as can be seen from (14), only the product forming the last term of) is of nth degree, and the factor of x^n then is 1. Therefore, $N^{(n)}(\xi) =$ $[x_0 \cdots x_n]$. That is,

) $$[x_0 \cdots x_n] = \frac{1}{n!} f^{(n)}(\xi), \qquad a \cdots \xi \cdots b.$$

e divided difference of nth order, multiplied by $n!$, is consequently ual to the nth derivative of the approximated function in an interval ich lies between the extreme values of $x_0, \cdots, x_n$. With the aid of is formula, the remainder term of Newton's formula may be put in a ferent form. If $\bar{\xi}$ is some value of the variable intermediate to the treme values of $x_0, \cdots, x_n$, then we may write, in place of equation 5),

$$R_{n+1} = \frac{(x - x_0) \cdots (x - x_n)}{(n + 1)!} f^{(n-1)}(\xi), \tag{25}$$

a form of the remainder which was first given by Cauchy.[3]

NOTES

1. *Principia* (1687), Book, III, Lemma V. case 2.
2. *Leçons élémentaires sur les mathématiques* (1795) Ouevres VII, p. 284.
3. Cauchy, *Comptes rendus Pa is* 11 (1840), pp. 775-789.

9. Calculation and Construction of Curves through given Points.

1. In interpolation, the problem is either to calculate, from the individual values of the function, which are sometimes given as points on a curve, the equation of the curve through these points, or to construct this curve. Various methods are used according to the type of data.

Mathematically, this problem is solved by the general interpolation formula, 8 (14).

If the function values are given for the arguments x_0 , x_1 , $\cdots$, x_n , then the constant y_0 is known. We first form the differences of first order,

$$[x_0x] = \frac{g(x) - g(x_0)}{x - x_0}$$

and substitute successively the values x_1 , $\cdots$, x_n ; $[x_0x_1]$ is then the coefficient of $(x - x_0)$. Next we form the second order divided differences from the above values. In turn we find the divided differences of third order, etc. This calculation of divided differences is best carried out in tabular form. In the execution of this scheme, it should be observed that in the formation of a divided difference of the next higher order, only differences of the preceding order are used. These have all the same values except one. The scheme is illustrated in the table below.

				$[x_nx_0] = \dfrac{y_n - y_0}{x_n - x_0}$			$[x_mx_1x_0] = \dfrac{[x_mx_0] - [x_1x_0]}{x_m - x_1}$
x_0	$\underline{y_0}$						
x_1	y_1	$x_1 - x_0$	$y_1 - y_0$	$\underline{[x_1x_0]}$			
x_2	y_2	$x_2 - x_0$	$y_2 - y_0$	$[x_2x_0]$	$x_2 - x_1$	$[x_2x_0] - [x_1x_0]$	$\underline{[x_2x_1x_0]}$
x_3	y_3	$x_3 - x_0$	$y_3 - y_0$	$[x_3x_0]$	$x_3 - x_1$	$[x_3x_0] - [x_1x_0]$	$[x_3x_1x_0]$
x_4	y_4	$x_4 - x_0$	$y_4 - y_0$	$[x_4x_0]$	$x_4 - x_1$	$[x_4x_0] - [x_1x_0]$	$[x_4x_1x_0]$
..	..	.....	.....	.....	.....	....	......

				$[x_r x_2 x_1 x_0] = \dfrac{[x_r x_1 x_0] - [x_2 x_1 x_0]}{x_r - x_2}$
x_0	$\underline{y_0}$			
x_1	y_1			
x_2	y_2			
x_3	y_3	$x_3 - x_2$	$[x_3 x_1 x_0] - [x_2 x_1 x_0]$	$\underline{[x_3 x_2 x_1 x_0]}$
x_4	y_4	$x_4 - x_2$	$[x_4 x_1 x_0] - [x_2 x_1 x_0]$	$[x_4 x_2 x_1 x_0]$
..	..	....	....	....

The underlined values are the values which occur as coefficients in the product development of $g(x)$.

2. *Example*: If we denote by γ' the number of moles of LiCl which are dissolved in 1000 gms of methyl alcohol, and by f_0 the osmotic coefficient, i.e., the ratio of the observed freezing point lowering to the decrease which is predicted by classical theory, we have, by the observations of Frivold:[1]

$2\gamma'$	0.074	0.148	0.232	0.354
$1 - f_0$	0.152	0.216	0.227	0.235

From these values we obtain the following table of divided differences, in which only the numbers above the horizontal line are to be considered.

$2\gamma'$	$1-f_0$									
0.074	$\underline{0.152}$									
0.148	0.216	0.074	0.064	$\underline{0.865}$						
0.232	0.227	0.158	0.075	0.475	0.084	−0.390	$\underline{-4.65}$			
0.354	0.235	0.280	0.083	0.296	0.206	−0.569	−2.76	0.122	1.89	15.5
0.341	0.230	0.267	0.078	0.293	0.193	−0.572	−2.96	0.109	1.69	15.5

$$1 - f_0 = 0.152 + 0.865\,(2\gamma' - 0.074) - 4.65\,(2\gamma' - 0.074)\,(2\gamma' - 0.148)$$
$$+ 15.5\,(2\gamma' - 0.074)\,(2\gamma' - 0.148)\,(2\gamma' - 0.232).$$

The expanded Horner method, given in 7.2 is used in the calculation of a function value. For example, as the value of $1 - f_0$ for $x_3 = 2\gamma' = 0.341$, we get

$x_3 - x$	0.109	0.193	0.267	
a_{n-1}	15.5	−4.65	+0.865	+0.152
$a_n(x_3 - x)$		+1.69	−0.572	+0.078
a'_{n-1}		−2.96	+0.293	+0.230.

This calculation can be carried out directly in the difference scheme, in which we start from the constant divided difference of third order. Then the numbers in the above table lying under the horizontal line are then obtained.

3. We consider once again the *construction of the curve by the method of Segner as generalized by Massau.* Here we shall limit ourselves to a function of third degree. Let the four points P_0, P_1, P_2, and P_3 be given,

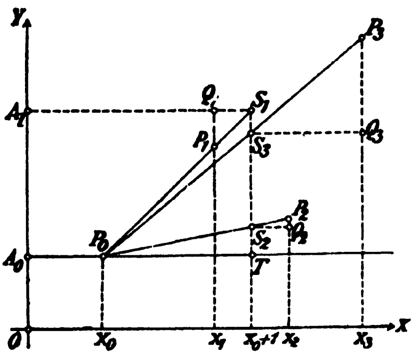


FIG. 29

(Fig. 29) and let us seek the coefficients a_0, a_1, a_2, and a_3. The value of the expression

$$y = a_0 + a_1(x - x_0) + a_2(x - x_0)(x - x_1) + a_3(x - x_0)(x - x_1)(x - x_2)$$

will be equal to a_0 for $x = x_0$, i.e., the x-parallel through P_0 cuts off the distance $0A_0 = a_0$ on the Y axis. First we determine the three points with abscissas x_1, x_2, and x_3, and ordinates $[x_0x_1]$, $[x_0x_2]$, $[x_0x_3]$, exactly as in the calculation scheme. Then P_0 is connected to P_1, P_2, and P_3. These connecting lines intercept the lengths

$$TS_1 = \frac{g(x_1) - g(x_0)}{x_1 - x_0} = [x_0x_1]; \qquad TS_2 = \frac{g(x_2) - g(x_0)}{x_2 - x_0} = [x_0x_2];$$

$$TS_3 = \frac{g(x_3) - g(x_0)}{x_3 - x_0} = [x_0x_3],$$

on the Y-parallel through $x_0 + 1$. If these points are projected on the Y-parallels passing through the corresponding abscissas (of P_1, P_2, P_3), we obtain the points Q_1, Q_2, and Q_3. These points have the desired coordinates if we consider a new x axis through A_0P_0. The projection of Q_1 on the Y axis gives the length $A_0A_1 = Q_1$. The same construction (which is not shown in Fig. 29) is carried on from Q_1 by use of the Y-parallel through $x_1 + 1$. We then obtain the points R_2, R_3, the ordinates of which, measured from the x-parallel through A_1Q_1, are calculated to be $[x_0x_1x_2]$ and $[x_0x_1x_3]$. Then $A_1A_2 = a_2$ is determined by R_2, and $A_2A_3 = a_3$ is also determined by the line R_2R_3 and the x-parallel through $x_2 + 1$. We proceed in the same fashion for equations of higher degree. From the values of the coefficients thus calculated, the curve on which lie the given points can be constructed point-by-point according to the method described in 7.9.

4. If we are concerned only with a pointwise construction of the interpolation curve, and not with the numerical determination of the coefficients, a simplification of the drawing can be undertaken. A special abscissa scale is introduced for each reduction step, as is shown in Fig. 30,

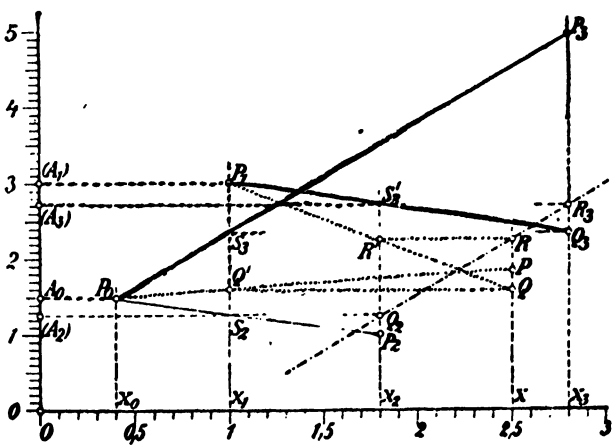


FIG. 30

where the interpolation curve is to be drawn through the four points $P_0(0.4, 1.5)$, $P_1(1, 3)$, $P_2(1.8, 1)$, $P_3(2.8, 5)$. The length $x_1 - x_0$ is chosen

as the unit for the first reduction, $x_2 - x_1$ for the second, etc. Then the Y-parallel at unit distance always passes through the next given point. Then this point remains unchanged in the reduction, so that Q_1 coincides with P_1, R_2 with Q_2, etc. In the construction of a new point then, the unit of measure for each interval must change correspondingly. In Fig. 30, the points P_1, S_2, S_3, and the corresponding abscissas P_1, Q_2, Q_3, are obtained from the given points P_0, P_1, P_2, P_3 by the reduction with the unit length $x_1 - x_0$. The unit for the second reduction is $x_2 - x_1$. By this reduction, the points Q_2 and R_3 are obtained, and consequently the line between them on which lie all points R. For example, the point R has the abscissa $x = 2.5$; if this point is projected by means of an x-parallel on the Y-parallel through $P_2(R')$, and if a line is drawn from P_1 through R back to the Y-parallel through R, the point Q is obtained, which lies on the curve $[xx_0]$. This point is projected on the Y-parallel through $P_1(Q')$, and the line through it from P_0 is drawn to the Y-parallel through R. This latter point is denoted by P. P lies on the interpolation curve. Dotted lines are used in the figure for all lines necessary for the construction of P. Thiele[3] has given a simple method for the construction of particular points of a curve of second order from these given points.

5. To carry out the determination of the coefficients of the interpolation curve by Lill's method, the methods of Art. 7 and 8 are combined. First, the given function values are measured from a point A_3 on the line on which the scale is to be constructed, and the points P_0, P_1, P_2, P_3 are obtained. A_3P_0 is then a_0. A perpendicular to the axis is then drawn through P_0 and lines are drawn through P_1, P_2, P_3 whose slopes (with respect to this perpendicular) are $x_1 - x_0$, $x_2 - x_0$, $x_3 - x_0$. These lines meet the normal in Q_1, Q_2, Q_3, and then

$$P_0Q_1 = [x_0x_1] = a_1 ; \qquad P_0Q_2 = [x_0x_2]; \qquad P_0Q_3 = [x_0x_3].$$

The perpendicular to the line P_0Q_1 is now erected at Q_1 and lines are drawn through Q_2 and Q_3 whose tangents with this line are $x_2 - x_1$ and $x_3 - x_1$ resp. The points of intersection of these lines then are

$$Q_1R_2 = [x_0x_1x_2] = a_2 ; \qquad Q_1R_3 = [x_0x_1x_3].$$

Finally, the perpendicular is drawn at R_2 and a line is drawn through R_3 such that the angle φ between this line and the perpendicular is given by $\operatorname{tg} \varphi = x_3 - x_2$. Then $OR_2 = [x_0x_1x_2x_3] = a_3$. The framework of the interpolation function now has been constructed. In the case shown in Fig. 31a, the same values are used as in Fig. 30. If further points are now to be constructed, e.g., $x = 2.5$, it is well first to make the accompanying drawing Fig. 31b, in order to obtain a more convenient determination of the proper directions. The direction $x - x_3$ is so determined that one

marks off from O_1 the distance x on OP_1, which is 2.5 in this case. The line connecting this point with the starting point S has the slope tg $\varphi = x - x_2$ with respect to SP_1, and gives the direction of the ray OR. Then $x = 2.5$ is plotted on O_2P_2, measuring from O_2. The line connecting this

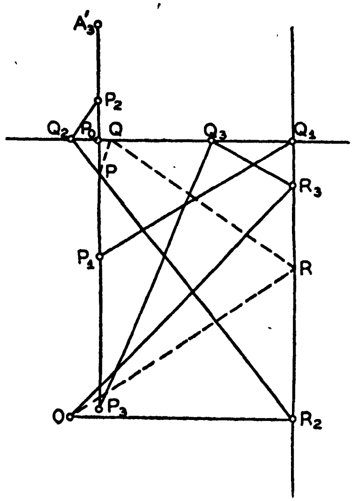


FIG. 31a

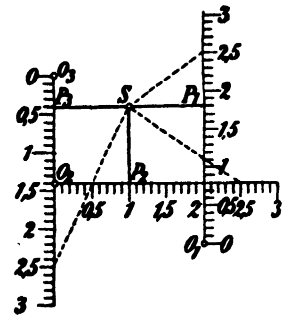


FIG. 31b

point with S has a tangent $x - x_2$ with SP_2, and this gives the slope of the second line RQ. By means of O_3P_3, the tangent of the third line QP can also be found. Then P is the point which corresponds to $x = 2.5$. These connecting lines are dotted in Fig. 31a.

6. If we have a scale in which the points lie so far apart, and in which the distances between successive points are so irregular that linear interpolation is no longer possible, then, according to the above construction, the scale can be very easily compressed with a quadratic interpolation. If the scale is to be compressed in the neighborhood of P_0, then the value $x = 0$ is assigned to P_0 and the values $+1$ and -1 are assigned to P_2 and P_1. On the perpendicular at P_0, the points Q_1 and Q_2 to which correspond the values $[x_1x_0]$ and $[x_1x_2]$, are obtained by reduction. The function $[x_1x]$ is linear. Therefore, to get the different polygon angles on this line segment, the interval from Q_1 and Q_2 is subdivided in equal parts. These points have the distances $a_1 + a_2(x + 1)$ from P_0. But the interpolation function is

$$y = a_1x + a_2x(x + 1).$$

The lines originating at these points must therefore have the tangents x with respect to Q_1P_0. To determine these slopes most simply, a uniform scale, with Q_1S as modulus, is drawn perpendicular to Q_1P_0 at S. The lines

from Q_1 to the various points then give the corresponding slopes. In Fig. 32, five points are constructed on either side. The other points of the interval will be constructed from P_1 or P_2, and consequently, a criterion exists as to whether or not a quadratic interpolation is sufficiently accurate in the degree of coincidences of the difference points.

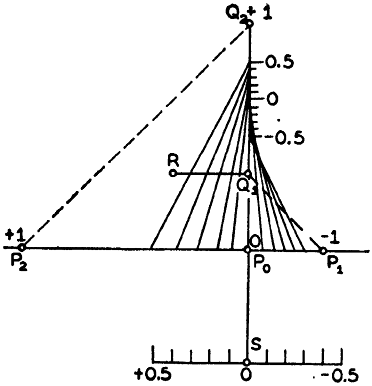


FIG. 32

NOTES

1. *Naturwissenschaften* 12 (1924), p. 414.
2. Thiele, *Interpolationsrechnung* (Leipzig, 1909), p. 148.

10. Special Interpolation Formulas.

1. We can set up the scheme given in 9.1 for the calculation of the divided differences from given values of the function, in a somewhat different way. We make use of the form given below. In this arrangement, the condition is fulfilled that two successive differences are always

				$\frac{y_m - y_{m-1}}{x_m - x_{m-1}}$		
x_0	y_0					
		$x_1 - x_0$	$y_1 - y_0$	$\underline{[x_1x_0]}$		
x_1	y_1				$x_2 - x_0$	$[x_2x_1] - [x_1x_0]$
		$x_2 - x_1$	$y_2 - y_1$	$[x_2x_1]$		
x_2	y_2				$x_3 - x_1$	$[x_3x_2] - [x_2x_1]$
		$x_3 - x_2$	$y_3 - y_2$	$[x_3x_2]$		
x_3	y_3				$x_4 - x_2$	$[x_4x_3] - [x_3x_2]$
..	..	....	....	....	....	....

		$\frac{[x_{m+1}x_m] - [x_mx_{m-1}]}{x_{m+1} - x_{m-1}}$			
x_0	y_0				
x_1	y_1	$\underline{[x_2x_1x_0]}$			
			$x_3 - x_0$	$[x_3x_2x_1] - [x_2x_1x_0]$	$\underline{[x_3x_2x_1x_0]}$
x_2	y_2	$[x_3x_2x_1]$			
			$x_4 - x_1$	$[x_4x_3x_2] - [x_3x_2x_1]$	$[x_4x_3x_2x_1]$
x_3	y_3	$[x_4x_3x_2]$			
..	..	....	....	....	....

identical in all argument values save one. Here also, the divided differences of higher order are formed from those of preceding order in the way previously described. Also, the divided difference in the top row of this table are the coefficients of the products in the equation of the interpolation curve 8 (14).

An example will be given here for the application of the general interpolation formula. This formula can be used to perform inverse interpolation in tables with equi-distant argument values, in which linear interpolation cannot be carried out. If we consider the inverse function of the tabulated function, this is naturally given for unequal argument intervals.

Example: The rate of interest is to be computed when, after 50 years, the principal has been increased by a factor of 5.15254693. From interest tables[1], we have

p	r^{50}
3.125	4.65798492
3.250	4.94883548
3.375	5.25746183
3.500	5.58492686
3.625	5.93235563

where p is the rate of interest and $r = 1 + p$. If the terms are arranged somewhat differently, so that we start from the value nearest to the given value, our scheme becomes

5.25746183	3.375			
		−0.30862635	−0.125	+0.405020505
4.94883548	3.250			
		+0.63609138	+0.250	+0.393025291
5.58492686	3.500			
		−0.92694194	−0.375	+0.404556083
4.65798492	3.125			
		+1.27437071	+0.500	+0.392350512
5.93235563	3.625			

+0.32746503	−0.011995214	−0.036630519			
			−0.59947691	−0.003014555	+0.005028642
−0.29085056	+0.011530792	−0.039645074			
			+0.98352015	+0.004513928	+0.004589363
+0.34742877	−0.012205571	−0.035131146			

+0.67489380	−0.000439279	−0.000650890.

Then

$$\begin{aligned} p = 3.375 &+ 0.405020505(x - x_0) - 0.036630519(x - x_0)(x - x_1) \\ &+ 0.005028642(x - x_0)(x - x_1)(x - x_2) \\ &- 0.000650890(x - x_0)(x - x_1)(x - x_2)(x - x_3), \end{aligned}$$

and in our case,

$x - x_0 = -0.10491490$; $x - x_1 = 0.20371145$; $x - x_2 = -0.43237993$; $x - x_3 = +0.49456201$.

If the general Horner scheme. given in 7.2, is used, then

−0.000650890	+0.005028642	−0.036630519	+0.405020505	+3.375
	−0.000321757	−0.002035167	−0.007876642	−0.04166631
	+0.004706885	−0.038665682	+0.397143863	+3.33333369 ≈ 3⅓%.

2. The calculations are especially simple if the values of the function are *equidistant.* This is generally the case if interpolation is performed directly in tables. If the function values $y_0, \cdots, y_n$ are arranged according to increasing argument, and if we set $x_1 = x_0 + h, \cdots$ or, more generally, $x_r = x_0 + rh$, then the increments of the argument in the formation of the divided differences of first order are all h, and in the differences of the second order are all $2h$, etc. In general, in the formation of the differences of the rth order, the increment is always rh. We have, therefore,

$$(1) \qquad [x_{m+1}x_m] = \frac{y_{m+1} - y_m}{h} = \frac{\Delta^1_{m+\frac{1}{2}}}{h}.$$

Here, $\Delta^1_{m+1/2}$ denotes the numerator difference, or more briefly, the *difference.* The upper index gives the order of the difference, while the subscript $m + 1/2$ determines the position of the difference in the series of the differences of first order. It is the arithmetic mean of the indices of the two values of the function from which the difference is formed. In a corresponding way, we have

$$(2) \qquad [x_{m+2}x_{m+1}x_m] = \frac{\Delta^1_{m+3/2} - \Delta^1_{m+1/2}}{h \cdot 2h} = \frac{\Delta^2_{m+1}}{2!\,h^2} = \frac{y_{m+2} - 2y_{m+1} + y_m}{2!h^2};$$

$$(3) \qquad [x_{m+3}x_{m+2}x_{m+1}x_m] = \frac{\Delta^2_{m+2} - \Delta^2_{m+1}}{h \cdot 2h \cdot 3h} = \frac{\Delta^3_{m+3/2}}{3!h^3} = \frac{y_{m+3} - 3y_{m+2} + 3y_{m+1} - y_m}{3!h^3}$$

and, in general, as can be shown by induction,

$$(4) \qquad [x_{m+l}x_{m+l-1} \cdots x_m] = \frac{\Delta^{l-1}_{m+(l+1)/2} - \Delta^{l-1}_{m+(l-1)/2}}{(l - 1)!h^{l-1} \cdot lh} = \frac{\Delta^l_{m+l/2}}{l!h^l}$$

$$= \frac{y_{m+l} - \binom{l}{1} y_{m+l-1} + \binom{l}{2} y_{m+l-2} \cdots \pm y_m}{l!h^l}.$$

According to section 8, equation (24) becomes

$$\Delta^l_{m+l/2} = h^l f^{(l)}(\xi), \quad \text{where} \quad x_m \leqq \xi \leqq x_{m+1} . \tag{5}$$

3. Since the divided difference of the lth order for equidistant ordinates is equal to the corresponding difference of the lth order, Δ^l, divided by $l!h^l$, the *difference scheme* can be simplified, and only the differences Δ^l need appear. We then obtain the system

x_0	y_0					
		$\Delta^1_{1/2}$				
$x_0 + 1h$	y_1		Δ^2_1			
		$\Delta^1_{3/2}$		$\Delta^3_{3/2}$		
$x_0 + 2h$	y_2		Δ^2_2		Δ^4_2	
		$\Delta^1_{5/2}$		$\Delta^3_{5/2}$		$\cdots$,
....	....	....	....	....	....	

where the difference is always written between the two values from which it is formed, so that each column begins a half row lower. The subscript of each difference makes it possible to determine the function values from which it is formed. For example,

$$\Delta^3_{5/2} = \Delta^2_3 - \Delta^2_2 = \Delta^1_{7/2} - 2\Delta^1_{5/2} + \Delta^1_{3/2} = y_4 - 3y_3 + 3y_2 - y_1$$

is composed of four values of the function. The indices of this value lie symmetrically about 5/2. In general, as was observed in equation (4),

$$\begin{aligned} \Delta^n_m &= y_{m+n/2} - \binom{n}{1} y_{m+n/2-1} + \binom{n}{2} y_{m+n/2-2} - \binom{n}{3} y_{m+n/2-3} \cdots \\ &\cdots (-1)^{n-1}\binom{n}{1} y_{m-n/2+1} + (-1)^n y_{m-n/2} = \sum_{r=0}^{r=n} (-1)^r \binom{n}{r} y_{m+n/2-r} . \end{aligned} \tag{6}$$

Since each term of the scheme is equal to the difference of the two preceding terms, any number in the scheme is equal to the sum of the terms of the succeeding column, from the first to the term which stands half a row above the term in question, plus the first term which occurs in the column of the number itself. For example,

$$\Delta^2_4 = \Delta^3_{7/2} + \Delta^3_{5/2} + \Delta^3_{3/2} + \Delta^2_1 = (\Delta^2_4 - \Delta^2_3) + (\Delta^2_3 - \Delta^2_2) + (\Delta^2_2 - \Delta^2_1) + \Delta^2_1 , \tag{7}$$

where all the terms cancel out except the first.

This property can be used as a *check on the difference scheme.* The sum of any column must equal the difference of the first and last terms in the preceding column.

Furthermore, each difference of the scheme can be calculated from the differences y_0 , $\Delta^1_{1/2}$, Δ^2_1 , $\Delta^3_{3/2}$, $\cdots$ occurring uppermost in the various columns. For example,

$$\Delta^1_{7/2} = \Delta^1_{5/2} + \Delta^2_3 = \Delta^1_{3/2} + 2\Delta^2_2 + \Delta^3_{5/2} = \Delta^1_{1/2} + 3\Delta^2_1 + 3\Delta^3_{3/2} + \Delta^4_2 .$$

More generally it can be shown that

$$\Delta^n_m = \sum_{r=0}^{m-n/2} \binom{m - n/2}{r} \Delta^{n+r}_{n+r/2} . \tag{8}$$

4. *Example.*[3] Various compressional forces are applied to a spring. The loads y in kg. and the corresponding contractions x in mm. are given in the following table:

mm	kg	Δ^1	Δ^2	Δ^3
0	0			
		48.7		
5	48.7		7.8	
		56.5		2.9
10	105.2		10.7	
		67.2		3.1
15	172.4		13.8	
		81.0		3.4
20	253.4		17.2	
		98.2		2.7
25	351.6		19.9	
		118.1		3.2
30	469.7		23.1	
		141.2		3.3
35	610.9		26.4	18.6
		167.6	118.9	
40	778.5	778.5		

The third difference oscillates about 3.1. These irregular fluctuations, which become larger in the succeeding differences, are to be attributed to errors in measurement. The sum of each column is given beneath it, and it can be seen that this sum is equal to the difference of the initial and final terms of the preceding column.

5. If the numbers of a table are not exact, but contain *errors of observation and approximation,* these errors will affect the differences, and consequently the result of the interpolation. If an error is made in the determination of the value of the argument of a function, it can be interpreted simply as the error of the value of the function. Then the function value determined by the inaccurate value of the argument is assigned to the correct value of the argument. The effects of such errors in the value

of the function will always be the same, whether the tabulated or observed function is to be found. Indeed, it can easily be seen from 8.2 that the effect of such an error on the differences will be as if the interpolation were performed in the table of errors, because the errors in the function values are additive. Moreover, it can be seen that the effect of the error of several functional values can be combined by addition of the effect of each individual error on the corresponding difference. If the difference scheme of the errors is so prepared that all the other errors are zero, and only the error, the effect of which is to be considered, is different from the zero, the following scheme is obtained:

	Δ^1	Δ^2	Δ^3	Δ^4
0		0		0
	0		0	
0		0		ϵ
	0		ϵ	
0		ϵ		-4ϵ
	ϵ		-3ϵ	
ϵ		-2ϵ		$+6\epsilon$
	$-\epsilon$		$+3\epsilon$	
0		ϵ		-4ϵ
	0		$-\epsilon$	
0		0		ϵ
	0		0	
0		0		0

The effects of such an error always spreads out, and is greatest in the row in which the error is made. The effect of the error on a column is determined by the magnitude of the binomial coefficients. However, these must be taken with alternating sign. This sometimes serves as a means for determining the error of calculation in a difference table.

In rounding off the values in a table, the error can amount to no more than one half unit in the last place. In the most unfavorable case, the following error table would be obtained. Here the errors are given in units of the last place:

	Δ^1	Δ^2	Δ^3	Δ^4
$+\frac{1}{2}$		-2		$+8$
	-1		$+4$	
$-\frac{1}{2}$		$+2$		-8
	$+1$		-4	
$+\frac{1}{2}$		-2		$+8$
	-1		$+4$	
$-\frac{1}{2}$		$+2$		-8
	$+1$		-4	
$+\frac{1}{2}$		-2		$+8$

The error of the rth difference is therefore at most $\pm 2^{r-1}$ units in the last place.

In a table prepared from observations, or in a table with approximated values, one should not expect that the differences of a certain order will disappear, even if a rational integral function is involved. Also, it should not be expected that the differences will fall below a given value, even if the interpolation series, which can be constructed for the function, is convergent. Therefore, in these cases—and we shall be most interested in such—there exists a behavior similar to that of divergent interpolation series. If the differences begin to oscillate irregularly about zero, the difference scheme should be terminated. We must then be content with an interpolation function of corresponding order. If the curve of the function is to be approximated by a rational integral function, we are limited, under such circumstances, to a stepwise representation of the function.

6. After these observations on the special difference table, we now consider the construction of the *interpolation formulas for equidistant function values.* If we start with the function value y_0, then the general interpolation formula of 8(14) takes the form

$$
\begin{aligned}
y = y_0 + \frac{x - x_0}{h} \frac{\Delta^1_{1/2}}{1!} + \frac{(x - x_0)(x - x_1)}{h^2} \frac{\Delta^2_1}{2!} + \cdots \\
+ \frac{(x - x_0) \cdots (x - x_{n-1})}{h^n} \frac{\Delta^n_{n/2}}{n!} + R_{n+1},
\end{aligned} \tag{9}
$$

where, by 8(25)

$$
R_{n+1} = \frac{(x - x_0)(x - x_1) \cdots (x - x_n)}{(n + 1)!} f^{(n+1)}(\xi), \tag{10}
$$

if ξ is a value intermediate to x_0 and $x_0 + nh$. We assume here that the interpolation formula is used for such an interval, i.e., that $x_0 \leqq x \leqq x_0 + nh$. This formula is usually known as *Newton's interpolation formula,* but it had already been mentioned by Gregory (without the remainder) in a letter to Collins on Nov. 23, 1670. In general, we are not concerned with the remainder. The formula is terminated at a suitably chosen difference. Another scale modulus is introduced for the abscissa for greater convenience in the calculation. We set $t = (x - x_0)/h$. Then, if we observe that

$$
(x - x_\lambda)/h = (x - x_0 - \lambda h)/h = t - \lambda,
$$

we obtain the Newton interpolation formula, without the remainder, from (9):

$$N_+(t) = y_0 + \frac{t}{1!}\Delta^1_{1/2} + \frac{t(t-1)}{2!}\Delta^2_1 + \frac{t(t-1)(t-2)}{3!}\Delta^3_{3/2} + \cdots$$

$$+ \frac{t(t-1)\cdots(t-n+1)}{n!}\Delta^n_{n/2} \tag{11}$$

$$= y_0 + \binom{t}{1}\Delta^1_{1/2} + \binom{t}{2}\Delta^2_1 + \binom{t}{3}\Delta^3_{3/2}\cdots\binom{t}{m}\Delta^m_{m/2} + \cdots + \binom{t}{n}\Delta^n_{n/2}.$$

The remainder takes the form

$$R_{n+1} = \binom{t}{n+1}h^{n+1}f^{(n+1)}(\tau), \tag{12}$$

where $0 \leqq \tau \leqq n$, if extrapolation is not performed.

Newton's formula is not useful for the calculation of the function values which are at any distance from the origin. For such values, the particular polynomials, with which the differences are to be multiplied, have very large values. Then the inaccuracies of the differences, multiplied by these large factors, enter into the calculated function values. In this way it happens that the differences of function values are determined which lie entirely on one side of the interval.

7. Just as differences which lie in descending lines in the difference scheme have been used in the above interpolation formula, a formula can also be constructed with the *differences* which lie in *ascending lines*, starting from the initial value of the function. To start again from the value y_0, we must elaborate the difference table above. For this purpose, we introduce negative indices:

$\cdots$					
x_{-3}	y_{-3}		Δ^2_{-3}		Δ^4_{-3}
		$\Delta^1_{-5/2}$		$\Delta^3_{-5/2}$	
x_{-2}	y_{-2}		Δ^2_{-2}		Δ^4_{-2}
		$\Delta^1_{-3/2}$		$\Delta^3_{-3/2}$	
x_{-1}	y_{-1}		Δ^2_{-1}		Δ^4_{-1}
		$\Delta^1_{-1/2}$		$\Delta^3_{-1/2}$	
x_0	y_0		Δ^2_0		Δ^4_0
		$\Delta^1_{+1/2}$		$\Delta^3_{+1/2}$	
x_{+1}	y_{+1}		Δ^2_{+1}		Δ^4_{+1}
		$\Delta^1_{+3/2}$		$\Delta^3_{+3/2}$	
x_{+2}	y_{+2}		Δ^2_{+2}		Δ^4_{+2}
		$\Delta^1_{+5/2}$		$\Delta^3_{+5/2}$	
x_{+3}	y_{+3}		Δ^2_{+3}		Δ^4_{+3} .
$\cdots$					

In the general interpolation formula, we must introduce successively x_0, $x_{-1} = x_0 - h$, $x_{-2} = x_0 - 2h$, etc., and obtain the formula

$$(13)\quad \begin{aligned} N_-(t) &= y_0 + \frac{t}{1!}\Delta^1_{-1/2} + \frac{t(t+1)}{2!}\Delta^2_{-1} + \frac{t(t+1)(t+2)}{3!}\Delta^3_{-3/2}\cdots \\ &\quad \cdots \frac{t(t+1)(t+2)\cdots(t+n-1)}{n!}\Delta^n_{-n/2} + R_{n+1}, \\ &= y_0 + \binom{t}{1}\Delta^1_{-1/2} + \binom{t+1}{2}\Delta^2_{-1} + \binom{t+2}{3}\Delta^3_{-3/2}\cdots \\ &\quad \cdots \binom{t+n-1}{n}\Delta^n_{-n/2} + R_{n+1}, \end{aligned}$$

after the introduction of the variable $t = (x - x_0)/h$. The remainder has the form

$$(14)\quad R_{n+1} = \frac{t(t+1)\cdots(t+n)}{(n+1)!}h^{n+1}f^{(n+1)}(\tau) = \binom{t+n}{n+1}h^{n+1}f^{(n+1)}(\tau)$$

and $-n \leqq \tau \leqq 0$, if the formula is not used for extrapolation; otherwise, the limits must be extended correspondingly. This formula is used in the numerical interpolation of differential equations (cf. Art. 32 and Art. 34). For actual interpolation, the formulas derived below are more commonly used.

8. It is more practical to begin with an intermediate value of the region to be represented, than with an end value as was used above. Also, it is better not to use differences which lie on a diagonal of the table, but only those which lie as close as possible to the initial value. This can be done by selecting as x_0 an abscissa in the middle of the given interval, and then substituting in the general interpolation formula:

$$(15)\quad x_0 = x_0,\ x_1 = x_0 - h,\ x_2 = x_0 + h,\ x_3 = x_0 - 2h,\ x_4 = x_0 + 2h\ \cdots.$$

The successive values of the function then lie about the first value, increasing on one side, and decreasing on the other by equal amounts. The interpolation formula is then

$$(16)\quad \begin{aligned} y &= y_0 + [x_{-1}x_0](x - x_0) + [x_{-1}x_0x_{+1}](x - x_0)(x - x_{-1}) \\ &\quad + [x_{-2}x_{-1}x_0x_1](x - x_0)(x - x_{-1})(x - x_{+1}) + \cdots + R_{n+1}, \end{aligned}$$

where, for odd n,

$$(17)\quad R_{n+1} = \frac{(x-x_0)(x-x_{-1})(x-x_{+1})\cdots(x-x_{-(n-1)/2})(x-x_{+(n-1)/2})f^{(n+1)}(\xi)}{(n+1)!}$$

in which $x_0 - (n-1)h/2 \leqq \xi \leqq x_0 + (n-1)h/2$. For even n, the last factors of the numerator are

$$(x - x_{-n/2+1})(x - x_{n/2-1})(x - x_{-n/2}).$$

This is the so-called *Gauss interpolation formula.* If the difference symbol is introduced in our special table, we have

$$\begin{aligned} y = y_0 + \Delta^1_{-1/2} \frac{(x - x_0)}{h} + \frac{\Delta^2_0}{2!} \frac{(x - x_0)(x - x_{-1})}{h^2} \\ + \frac{\Delta^3_{-1/2}}{3!} \frac{(x - x_0)(x - x_{-1})(x - x_{+1})}{h^3} \cdots \end{aligned} \tag{18}$$

and if the variable $t = (x - x_0)/h$ is again introduced,

$$\begin{aligned} G_1(t) &= y_0 + \Delta^1_{-1/2}t + \frac{t(t+1)}{2!}\Delta^2_0 + \frac{t(t^2-1)}{3!}\Delta^3_{-1/2} \\ &\quad + \frac{t(t^2-1)(t+2)}{4!}\Delta^4_0 \cdots \\ &= y_0 + \binom{t}{1}\Delta^1_{-1/2} + \binom{t+1}{2}\Delta^2_0 + \binom{t+1}{3}\Delta^3_{-1/2} \\ &\quad + \binom{t+2}{4}\Delta^4_0 + \binom{t+2}{5}\Delta^5_{-1/2} \cdots, \end{aligned} \tag{19}$$

where the remainder is

$$R_{n+1} = \binom{t + (n+1)/2}{n+1} f^{(n+1)}(\tau) \tag{20}$$

for odd n, and

$$R_{n+1} = \binom{t + n/2}{n+1} f^{(n+1)}(\tau) \tag{21}$$

for even n. The value τ is intermediate to the limits of the argument. In this formula, the differences are used which lie in the row of y_0 and half a row higher. In the same manner, a formula may be constructed which uses the differences of the row of y_0, and those which lie a half row lower. This formula is obtained if the argument values x_0, $x_0 + h$, $x_0 - h$, $x_0 + 2h$, $x_0 - 2h$, $\cdots$ are introduced in the general interpolation formula 8(14). After the introduction of the variable t, we have

$$G_2(t) = y_0 + t\Delta^1_{1/2} + \frac{t(t-1)}{2!}\Delta^2_0 + \frac{t(t^2-1)}{3!}\Delta^3_{1/2} \tag{22}$$

$$+ \frac{t(t^2 - 1)(t - 2)}{4!} \Delta_0^4 \cdots$$

$$G_2(t) = y_0 + \binom{t}{1} \Delta_{1/2}^1 + \binom{t}{2} \Delta_0^2 + \binom{t+1}{3} \Delta_{1/2}^3 \tag{22}$$

$$+ \binom{t+1}{4} \Delta_0^4 + \binom{t+2}{5} \Delta_{1/2}^5 \cdots$$

with the remainder (21) for even n, while for odd n,

$$R_{n+1} = \binom{t + (n-1)/2}{n+1} f^{(n+1)}(\tau).$$

9. Since these formulas exhibit a certain asymmetry, the mean has been formed. The formula which is then obtained is known as *Stirling's interpolation formula*[3]

$$S(t) = y_0 + t\,\frac{\Delta_{-1/2}^1 + t_{+1/2}^1}{2} + \frac{t^2}{2!}\Delta_0^2 + \frac{t(t^2-1)}{3!}\,\frac{\Delta_{-1/2}^3 + \Delta_{+1/2}^3}{2}$$

$$+ \frac{t^2(t^2-1)}{4!}\Delta_0^4 \cdots \tag{23}$$

$$= y_0 + t\overline{\Delta_0^1} + \frac{t^2}{2!}\Delta_0^2 + \frac{t(t^2-1)}{3!}\overline{\Delta_0^3} + \frac{t^2(t^2-1)}{4!}\Delta_0^4 \cdots,$$

where we write

$$\overline{\Delta}_0^{2\kappa-1} = \frac{\Delta_{1/2}^{2\kappa-1} + \Delta_{-1/2}^{2\kappa-1}}{2} \tag{23a}$$

for the mean of two successive differences. The remainder is equal to the arithmetic mean of the remainders of the formula (19) and (22). Stirling's formula is especially useful in the neighborhood of the value y_0, because in the immediate vicinity of y_0, the values of the polynomial are small and consequently the effect of higher differences and their errors are small.

10. The formulas which have been constructed can be used to increase the number of values in a table; i.e., to insert *other values* between the between the *given values*. This interpolation is convenient only when the tabulated values lie so close together that the curve connecting them can be approximated by a straight line; i.e., whenever the second differences are so small that they can be neglected. Then we have linear interpolation, as was discussed in Art. 4. To simplify the calculation in the general case, tables are used for the polynomials with which the particular differences

are to be multiplied, and it is customary to subdivide the interval into ten or one hundred equal parts. All the formulas developed so far are especially advantageous for interpolation about a point. Because of this, such tables are usually given between the limits -0.5 and $+0.5$.[4]

11. If, in addition to G_2, we construct a formula analogous to G_1, which starts from y_1 instead of y_0, and which therefore uses the differences of the columns 1 and 1/2, we obtain

$$G_3(t) = y_1 + \Delta^1_{1/2}(t-1) + \Delta^2_1 \frac{(t-1)t}{2!} + \Delta^3_{1/2}\frac{(t-1)t(t-2)}{3!} + \Delta^4_1 \frac{t(t-2)(t^2-1)}{4!} + \cdots$$

(24)

$$= y_1 + \Delta^1_{1/2}\binom{t-1}{1} + \Delta^2_1\binom{t}{2} + \Delta^3_{1/2}\binom{t}{3} + \Delta^4_1\binom{t+1}{4} + \Delta^5_{1/2}\binom{t+1}{5}\cdots,$$

with the remainder

$$R_{n+1} = \binom{t+(n-1)/2}{n+1} f^{(n+1)}(\tau) \tag{25}$$

for odd n, and

$$R_{n+1} = \binom{t+(n-2)/2}{n+1} f^{(n+1)}(\tau) \tag{26}$$

for even n. Here also, τ is a value between the extreme values. If we form the mean of G_2 and G_3, then we have *Bessel's interpolation formula.* The differences of odd order common to both formulas lie in the middle of the interval 0 to 1, while the various differences of even order lie at the limits. The formula becomes

$$B(t) = \frac{y_0+y_1}{2} + \Delta^1_{1/2}\left(t-\frac{1}{2}\right) + \frac{\Delta^2_0+\Delta^2_1}{2}\frac{t(t-1)}{2!}$$

(27)

$$+ \Delta^3_{1/2}\frac{t(t-1)(t-1/2)}{3!} + \frac{\Delta^4_0+\Delta^4_1}{2}\frac{t(t-2)(t^2-1)}{4!}\cdots,$$

Since $\Delta^1_{1/2} = y_1 - y_0$, $(y_0 + y_1)/2 + \Delta^1_{1/2}(t - 1/2) = y_0 + \Delta^1_{1/2}$, and the Bessel formula may also be written in the form

$$B(t) = y_0 + \Delta^1_{1/2}t + \overline{\Delta^2_{1/2}}\frac{t(t-1)}{2!} + \Delta^3_{1/2}\frac{t(t-1)(t-1/2)}{3!} + \overline{\Delta^4_{1/2}}\frac{t(t-2)(t^2-1)}{4!}\cdots, \tag{28}$$

where the barred terms represent the mean values. The remainder is again the arithmetic mean of the remainders of formulas (22) and (24). The factor $(t - 1/2)$ enters into the formula from the formation of the mean value. Bessel's formula is not used for interpolation in the vicinity of a point, but for interpolation throughout an entire interval. Therefore, tables are given for values of the polynomial from 0 to 1.[5]

Occasionally $t = \tau + 1/2$ is introduced as a new variable. Then the first form of Bessel's formula (the more convenient form) becomes

$$B(\tau + 1/2) = \frac{y_0 + y_1}{2} + \Delta^1_{1/2}\tau + \frac{\overline{\Delta^2_{1/2}}}{2!}\left(\tau^2 - \frac{1}{4}\right) + \frac{\Delta^3_{1/2}}{3!}\tau\left(\tau^2 - \frac{1}{4}\right) + \frac{\overline{\Delta^4_{1/2}}}{4!}\left(\tau^2 - \frac{1}{4}\right)\left(\tau^2 - \frac{9}{4}\right) + \cdots, \tag{29}$$

where the bars again indicate mean values.

12. *Example*: The value of the principal and interest accruing on one dollar at 3 1/3% for 50 years is to be computed.[6] From interest tables we have

x	t	r^n	Δ^1	Δ^2	Δ^3	Δ^4
3.125	−2	4.65798492				
			0.29085056			
3.250	−1	4.94883548		0.01777579		
		5.103148655	0.30862635	*0.018307235*	0.00106289	
3.375	0	5.25746183	*0.31804569*	0.01883868	*0.001093975*	0.00006217
			0.32746503		0.00112506	
3.500	+1	5.58492686		0.01996374		
			0.34742877			
3.625	+2	5.93235563				

If the fourth difference is assumed to be constant, it follows from Bessel's formula (if we start from $x = 3.250$ and set $t = 2/3$),

$$y = 5.103148655 + 0.30862635(1/6) + 0.018307235\,\frac{2/3 \times (-1/3)}{1\cdot 2}$$

$$+ 0.00106289 \frac{2/3 \times (-1/3) \times 1/6}{1\cdot 2\cdot 3}$$

$$+ 0.00006217 \frac{2/3 \times (-4/3) \times (-5/9)}{1\cdot 2\cdot 3\cdot 4}$$

$$= 5.103148655 + 0.051437725 - 0.0020341372 - 0.0000065610$$
$$+ 0.000001279$$
$$= 5.15254696.$$

By Stirling's formula, we have for $t = -1/3$,

$$y = 5.25746183 - \left(\frac{1}{3}\right) 0.31804569 + \left(\frac{1}{18}\right) 0.01883868$$
$$+ \left(\frac{4}{81}\right) 0.001093975 - \left(\frac{1}{243}\right) 0.00006217$$
$$= 5.15254695.$$

If we were to start from $x = 3.125$ and use the Newtonian formula N_+, or use N_- starting from 3.625, we would have obtained the same values only that the factors with which the differences would be multiplied would have been larger. But other values could be obtained if, starting from 3.250 or 3.375 we used either the formula N_- or N_+, since the fourth differences are not then constant.

13. For interpolation in an interval, a form of the Gaussian formula, due to Laplace,[7] can be given, in which only those differences appear which occur in the series of both values between which one is interpolating. If we substitute the even differences of next lower order in

$$G_2(t) = y_0 + \binom{t}{1} \Delta^1_{1/2} + \binom{t}{2} \Delta^2_0 + \binom{t+1}{3} \Delta^3_{1/2}$$
(22)
$$+ \binom{t+1}{4} \Delta^4_0 + \binom{t+2}{5} \Delta^5_{1/2} \cdots$$

for the differences of odd order from which they are formed, we have

$$G_2(t) = y_0 + \binom{t}{1} (y_1 - y_0) + \binom{t}{2} \Delta^2_0 + \binom{t+1}{3} (\Delta^2_1 - \Delta^2_0)$$
(22a)
$$+ \binom{t+1}{4} \Delta^4_0 + \binom{t+2}{5} (\Delta^4_1 - \Delta^4_0) \cdots .$$

If we further observe the well-known relation between binomial coefficients

$$\binom{t+1}{r+1} = \binom{t}{r+1} + \binom{t}{r},$$

we can collect the factors of each difference and obtain

$$\begin{aligned} G_2(t) &= (1-t)y_0 + y_1 \cdot t - \binom{t}{3}\Delta_0^2 + \binom{t+1}{3}\Delta_1^2 \\ &\qquad - \binom{t+1}{5}\Delta_0^4 + \binom{t+2}{5}\Delta_1^4 \cdots, \end{aligned} \tag{30}$$

or, in a somewhat different arrangement,

$$\begin{aligned} G_2(t) &= (1-t)y_0 - \binom{t}{3}\Delta_0^2 - \binom{t+1}{5}\Delta_0^4 + \cdots \\ &\qquad + ty_1 + \binom{t+1}{3}\Delta_1^2 + \binom{t+2}{5}\Delta_1^4 \cdots . \end{aligned} \tag{31}$$

If we set $1 - t = \tau$ we obtain the formula given by *Laplace*[8]

$$\begin{aligned} B(t) &= \tau \cdot y_0 + \binom{\tau+1}{3}\Delta_0^2 + \binom{\tau+2}{5}\Delta_0^4 + \cdots \\ &\qquad + ty_1 + \binom{t+1}{3}\Delta_1^2 + \binom{t+2}{5}\Delta_1^4 \cdots . \end{aligned} \tag{32}$$

Tables can be used here also for the coefficients.

14. Ordinarily, one is not further interested in the *remainder term* when working with interpolation series. It is assumed that the term is small enough to be neglected. Nevertheless, it is advisable to observe the magnitude of the first difference which is neglected. This gives an approximate measure of the magnitude of the remainder term, and gives the computer a general idea of the magnitude of the formula error. The error caused by neglecting the remainder term is not always small throughout the entire interval. But if the argument intervals of the scheme are so small that only a few terms of the interpolation formula are needed, and if we are satisfied with a limited accuracy, then we can in general

get a good approximation. However, if we want to extend the interpolation accuracy over a certain interval, in the case of a non-vanishing remainder term, for rather large intervals, we can obtain incorrect results, because the remainder term should not be neglected. We shall omit here any investigation of convergence, by which it can be proved that the interpolation series are frequently semi-convergent series. It could also be shown that the convergence rates are better for the Newton series, used for practical calculations only within the limits of the tables, than for any other series, although in these latter series, the first terms give a better approximation.[9] We are interested in interpolation series from the viewpoint of practical calculations; the convergence itself is not important. We are concerned with the inaccuracy, and this is to be neglected if the remainder term is small after a few terms of the series.

15. In conclusion, we will consider an application of interpolation series.[10] In the statistical analysis of fluctuating processes, we determine the number of individual cases which lie in a certain interval of variation of the fluctuating argument. If, for example, we want to determine the relative heights of a group of people, then we determine the number of people whose heights lie between 140 and 145 cm., between 145 and 150 cm., etc. If these numbers are available, the question arises: how many people have a height which lies between narrower limits, e.g., between 162 and 163 cm.

If $y_0, y_1, \cdots$ are the values of the function between the desired smaller limits and $2r + 1$ of these values, taken together, give the value in the table for the larger interval, then

$$(33)\quad Y_{-1} = \sum_{-(3r+1)}^{-(r+1)} y_m;\ Y_0 = \sum_{-r}^{+r} y_m;\ Y_1 = \sum_{r+1}^{3r+1} y_m;\ Y_2 = \sum_{3r+2}^{5r+2} y_m \cdots .$$

Then we can introduce a function

$$(34)\qquad z_x = \sum_{m=p}^{(2r+1)x-r-1} y_m$$

in order to calculate a value y_0 lying in the middle of a large interval. Here p is a completely arbitrary number. By means of this function, we can express the values lying in the table as differences of the first order:

$$(35)\qquad Y_n = \Delta^1_{n+1/2} .$$

Furthermore, y_0 can also be expressed as a function of the z's:

$$(36)\qquad y_0 = z_{1/2+1/(2(2r+1))} - z_{1/2-1/(2(2r+1))} .$$

If we now introduce the values $\tau = \pm 1/(2(2r + 1))$ in Bessel's formula (29), we get

(37)
$$B_{1/2}\left(\frac{1}{2} \pm \frac{1}{2(2r+1)}\right) = \bar{y}_{1/2} \pm \Delta^1_{1/2}\tau + \overline{\Delta^2_{1/2}}\left(\frac{\tau^2 - 1/4}{2!}\right) \pm \Delta^3_{1/2}\frac{\tau(\tau^2 - 1/4)}{3!} + \overline{\Delta^4_{1/2}}\frac{(\tau^2 - 1/4)(\tau^2 - 9/4)}{4!} \cdots,$$

and if we form the difference, we have

(38)
$$B\left(\frac{1}{2} + \frac{1}{2(2r+1)}\right) - B\left(\frac{1}{2} - \frac{1}{2(2r+1)}\right) = 2\tau\Delta^1_{1/2} + \frac{2\tau(\tau^2 - 1/4)}{3!}\Delta^3_{1/2} + \frac{2\tau(\tau^2 - 1/4)(\tau^2 - 9/4)}{5!}\Delta^5_{1/2}.$$

If we recall the equations (35) and (36), it follows that

(39)
$$y_0 = 2\tau Y_0 + \frac{2\tau(\tau^2 - 1/4)}{3!}\Delta^2_0 Y + \frac{2\tau(\tau^2 - 1/4)(\tau^2 - 9/4)}{5!}\Delta^4_0 Y \cdots.$$

If, for example, we assume that 5 values of y are given together in Y, then, by (33), $2r + 1 = 5$, and therefore $\tau = 0.1$. From (39) we get

(40)
$$y_0 = 0.2Y_0 - 0.008\Delta^2_0 Y + 0.000896\Delta^4_0 Y \cdots.$$

Example: According to the German mortality table, of 100,000 people who reach their 20th year, the number of people Y is given who die in the various 5 year intervals following that age:

	Y	Δ^1	Δ^2	Δ^3	Δ^4
20—25	4040				
		72			
25—30	4112		267		
		339		+ 71	
30—35	4451		338		−306
		677		−235	
35—40	5128		103		
		780			
40—45	5908				

By the above formula therefore, in the 33rd year $y_0 = 0.2 \times 4451 - 0.008 \times 338 + 0.000896 \times (-306) = 887.2$ people died. Actually, the table gives the value 888.

Naturally, such good results are not obtained for arbitrary values. For example, we consider the coal production in Prussia; this amounted (in thousands of tons) to the values in the following table:

	Y	Δ^1	Δ^2	Δ^3	Δ^4
1897—1901	471740				
		+ 90537			
1902—1906	562277		+ 55234		
		+145771		−118641	
1907—1911	708048		− 63407		−7756.
		+ 82364		−126397	
1912—1916	790412		−189804		
		−107440			
1917—1921	682972				

Consequently the production would have been $y_0 = 0.2 \times 708048 + 0.008 \times 63407 - 0.000896 \times 7756 = 142110$ in 1909. In reality, not 142,110,000 tons but 139,906,000 tons were mined. The error is therefore 1 1/2%.

NOTES

1. Förster, *Politische Arithmetik* (Leipzig, 1924), p. 33. Spitzer, *Tabellen für die Zinsezinsen-und Rentenrechnung*, 6th ed. (Vienna, 1922).
2. From an example by R. Rothe in his lectures.
3. Newton, *Methodus differentialis* (1711), Prop. III, Case 1; Stirling, *Methodus differentialis* (1730), Prop. XX.
4. For example, Bruns, *Grundlinien des wissenschaftlichen Rechnens* (Leipzig, 1903), pp. 43-44. Lindow, *Numerische Infinitesimalrechnung* (Berlin and Bonn, 1928), p. 158 ff.
5. V. note 4.
6. Förster, *Politische Arithmetik* (Leipzig, 1924), p. 21.
7. Laplace, *Théorie analytique des probabilités* (Paris, 1812).
8. This formula is also known as Everett's formula. *Brit. Assoc. Rep.* (1900).
9. Nörlund, *Differenzenrechnung* (Berlin, 1924), Ch. 8.
10. King, *Journal of the Institute of Actuaries*, 43 (1909), p. 114.

11. Numerical Differentiation.

1. Up to now, in order to avoid ambiguities, we have assumed not only that the function values entering into the difference scheme are single-valued and finite, but also that the argument values are all different. We shall now remove this last assumption. For example, if any two argument values x_r and x_m are equal, then the divided difference of the $(r - m)$th order

$$(1)\quad [x_m x_{m+1} \cdots x_{r-1} x_r] = \frac{[x_{m+1} x_{m+2} \cdots x_{r-1} x_r] - [x_m x_{m+1} \cdots x_{r-1}]}{x_r - x_m}$$

is indeterminate. If $f(x)$ possesses a derivative at x_m, then this expression tends to a limit for $x_r \to x_m$, and indeed, we have

$$\lim_{x_r \to x_m} [x_r x_m x_{m+1} \cdots x_{r-1}] = [x_m x_m x_{m+1} \cdots x_{r-1}] \tag{2}$$

$$= \frac{d}{dx} [x_m x_{m+1} \cdots x_{r-1}].$$

This expression is known as the *divided difference with repeated argument*. That the condition: $f'(x)$ exists for $x = x_m$ is sufficient in order that the divided difference with repeated argument have a determinate value, follows from the considerations below. According to 8(10),

$$\begin{aligned}
[xx_m \cdots x_{r-1}] \\
&= \frac{f(x)}{(x - x_m)(x - x_{m+1}) \cdots (x - x_{r-1})} \\
&\quad + \frac{f(x_m)}{(x_m - x)(x_m - x_{m+1}) \cdots (x_m - x_{r-1})} \\
&\quad + \frac{f(x_{m+1})}{(x_{m+1} - x)(x_{m+1} - x_m)(x_{m+1} - x_{m+2}) \cdots (x_{m+1} - x_{r-1})} \cdots \\
&\quad \cdots \frac{f(x_{r-1})}{(x_{r-1} - x)(x_{r-1} - x_m) \cdots (x_{r-1} - x_{r-2})}.
\end{aligned} \tag{2a}$$

If x is allowed to approach x_m, then the terms from the third on all have a finite value. The first two terms are not defined for $x = x_m$. We must therefore find their limit. Since the limit of a product equals the product of the limits of the factors,

$$\begin{aligned}
&\lim_{x \to x_m} \frac{f(x)}{(x - x_m)(x - x_{m+1}) \cdots (x - x_{r-1})} \\
&\qquad - \lim_{x \to x_m} \frac{f(x_m)}{(x - x_m)(x_m - x_{m+1}) \cdots (x_m - x_{r-1})} \\
&\qquad = \frac{1}{(x_m - x_{m+1}) \cdots (x_m - x_{r-1})} \lim_{x \to x_m} \frac{f(x) - f(x_m)}{x - x_m} \\
&\qquad = \frac{f'(x_m)}{(x_m - x_{m+1}) \cdots (x_m - x_{r-1})}.
\end{aligned} \tag{3}$$

In particular,

$$[xx] = f'(x). \tag{4}$$

We can also form divided differences with manifold (say l) repetitions of the argument. These differences will exist if the first l derivatives of $f(x)$ exist. It should be mentioned here that

$$\underset{n}{[x \cdots x]} = \frac{1}{n!} f^{(n)}(x). \tag{5}$$

This follows directly from 8(24), where the relation between derivatives and divided differences is given. If we let the $(n + 1)$ values $x_0, \cdots, x_n$ approach x, then the intermediate value ξ also converges to x, since ξ lies between the largest and the smallest of the values $x_0, \cdots, x_n$. The general Newton interpolation formula goes over into Taylor's formula with the use of these divided differences with repeated argument.

2. An *approximate value for the derivative of a tabulated function* can be formed on the basis of this connection between divided differences with repeated argument and derivatives. Naturally, we can also find approximate values by termwise differentiation of the general Newton interpolation formula.[1] Consider a table of the cubes of the prime numbers; i. e., a table of the function $y = x^3$, with varying argument intervals. To find the derivative for $x = 9$, we make use of the table below, forming according to the scheme of 10.1:

1	1									
		1	7	7						
2	8				2	12	6			
		1	19	19				4	4	1
3	27				3	30	10			
		2	98	49				5	5	1
5	125				4	60	15			
		2	218	109				8	8	1
7	343				6	138	23			
		4	988	247				8	8	1
11	1331				6	186	31			
		2	866	433				2	2	1
13	2197				-2	-66	33			
		-4	-1468	367				-2	-2	1
9	729				-4	-124	31			
		0	0	$\underline{\underline{243}}$				-4	-4	1
9	729				0	0	$\underline{\underline{27}}$			
		0	0	243				0	0	$\underline{\underline{1}}$
9	729				0	0	27			
		0	0	243				$x-9$	$x-9$	1
9	729				$x-9$	$x^2+9x-162$	$x+18$			
		$x-9$	x^3-729	$x^2+9x+81$				$x-9$	$x-9$	1
x	x^3				$x-9$	$2x^2-9x-81$	$2x+9$			
		0	0	$\underline{\underline{3x^2}}$				$x-9$	$x-9$	1
x	x^3				0	0	$\underline{\underline{3x}}$			
		0	0	$3x^2$				$x-9$	$x-9$	$\underline{\underline{1}}$
x	x^3									

In the representation, the scheme is carried out to the constant third difference, and is further built up beyond $x = 13$. From the known argu-

ment differences and the known divided differences, we can form the differences of the divided differences of the preceding column. It can be seen from this arrangement that the divided differences with repeated argument can be calculated very simply. The underlined values give, for $x = 9$, $y' = [xx] = 243$, $y'' = 2[xxx] = 54$, $y''' = 6[xxxx] = 6$. Furthermore, x is then introduced as the argument. The underlined values give the derivatives, except for the factors 1!, 2!, 3!

3. *Example*: We seek to calculate the derivative $y' = (1 - f)'$ for $x = 0.341$ in the example given in 9.2. For that purpose, we repeat the table here, but now use the scheme 10.1 and carry it on with the repeated argument $x = 0.341$:

0.074	0.152									
		0.024	0.064	0.865						
0.148	0.216				0.158	−0.735	−4.65			
		0.084	0.011	0.131				0.280	4.31	15.5
0.232	0.227				0.206	−0.070	−0.340			
		0.122	0.008	0.061				0.193	2.99	15.5
0.354	0.235				0.109	+0.289	+2.65			
		−0.013	−0.005	0.350				0.109	1.69	15.5
0.341	0.230				−0.013	−0.057	+4.34			
		0	0	0.293						
0.341	0.230									

Therefore the rate of change of the osmotic coefficient with the concentration is $f' = -0.293$ at $2\gamma = 0.341$. If we can terminate the difference scheme with divided differences of a comparatively low order, the calculation of a difference quotient by means of the divided differences with repeated argument is comparatively simple. But it must be assumed that not only the function to be approximated, but also the derivatives of that function are well represented by the interpolation function in the neighborhood of the argument values in question.

4. If a difference table has already been set up for *equidistant values of the argument*, then we can start from the representation of the function by special interpolation formulas and differentiate these termwise. This is permissible if the function and its derivatives are approximated with sufficient accuracy with a few terms. In general, only Stirling's and Bessel's formulas are used. We shall also use Newton's formula N_-, but only in 32.8. There the corresponding values are also given. We recall that we have introduced the variable $t = (x - x_0)/h$, that, therefore, $df/dx = df/dt \cdot dt/dx = df/dt \cdot 1/h$ so that $df/dt = hf'(x)$ and if we write $\Delta_0^m = (\Delta_{-1/2}^m + \Delta_{1/2}^m)/2$ for the mean of two successive differences, we obtain for the first two differences, from Stirling's formula 10(23):

$$
\begin{aligned}
\frac{dS}{dt} &= hS'(x) = \overline{\Delta_0^1} \\
&\quad + \Delta_0^2 t + \overline{\Delta_0^3}\,\frac{3t^2 - 1}{3!} + \Delta_0^4\,\frac{4t^3 - 2t}{4!} + \overline{\Delta_0^5}\,\frac{5t^4 - 15t^2 + 4}{5!} \\
&\quad + \Delta_0^6\,\frac{6t^5 - 20t^3 + 8t}{6!} + \overline{\Delta_0^7}\,\frac{7t^6 - 70t^4 + 147t^2 - 36}{7!} + \cdots, \\
\frac{d^2S}{dt^2} &= h^2S''(x) = \Delta_0^2 \\
&\quad + \overline{\Delta_0^3}\,t + \Delta_0^4\,\frac{12t^2 - 2}{4!} + \overline{\Delta_0^5}\,\frac{20t^3 - 30t}{5!} \\
&\quad + \Delta_0^6\,\frac{30t^4 - 60t^2 + 8}{6!} + \overline{\Delta_0^7}\,\frac{42t^5 - 280t^3 + 294t}{7!} + \cdots,
\end{aligned}
\tag{6}
$$

and similarly, from Bessel's formula 10(27):

$$
\begin{aligned}
\frac{dB}{dt} &= hB'(x) = \Delta_{1/2}^1 + \overline{\Delta_{1/2}^2}\,\frac{2t - 1}{2} + \Delta_{1/2}^3\,\frac{6t^2 - 6t + 1}{12} \\
&\quad + \overline{\Delta_{1/2}^4}\,\frac{2t^3 - 3t^2 - t + 1}{12} + \Delta_{1/2}^5\,\frac{5t^4 - 10t^3 + 5t - 1}{120} + \cdots, \\
\frac{d^2B}{dt^2} &= h^2B''(x) = \overline{\Delta_{1/2}^2} + \Delta_{1/2}^3\,\frac{2t - 1}{2} + \overline{\Delta_{1/2}^4}\,\frac{6t^2 - 6t - 1}{12} \\
&\quad + \Delta_{1/2}^5\,\frac{4t^3 - 6t^2 + 1}{24} + \cdots.
\end{aligned}
\tag{7}
$$

Tables have also been prepared for these polynomials.[2] In general it is so arranged that the values of the derivatives are formed from the first group for the argument values from which the interpolation formula has been constructed. Therefore t is set equal to zero in equation (6):

$$
\begin{aligned}
S'(0) &= hS'(x_0) = \overline{\Delta_0^1} - \frac{1}{6}\,\overline{\Delta_0^3} + \frac{1}{30}\,\overline{\Delta_0^5} - \frac{1}{140}\,\overline{\Delta_0^7} \cdots \\
S''(0) &= h^2S''(x_0) = \Delta_0^2 - \frac{1}{12}\,\Delta_0^4 + \frac{1}{90}\,\Delta_0^6 \cdots.
\end{aligned}
\tag{8}
$$

In equations (7), which are derived from the Bessel interpolation series (and which are especially good approximations in the middle of the interval), we set $t = 1/2$ and obtain the derivatives in the middle of the interval

$$
(9)\qquad
\begin{aligned}
hB'\left(x_0 + \frac{h}{2}\right) &= \Delta^1_{1/2} - \frac{1}{24}\,\Delta^3_{1/2} + \frac{3}{640}\,\Delta^5_{1/2} \cdots ,\\
h^2B''\left(x_0 + \frac{h}{2}\right) &= \overline{\Delta}^2_{1/2} - \frac{5}{24}\,\overline{\Delta}^4_{1/2} + \frac{259}{5760}\,\overline{\Delta}^6_{1/2} \cdots .
\end{aligned}
$$

If we have the derivatives for other values, we can form a difference scheme for the values calculated, and then interpolate the desired values.

In the formulas S'' and B', differences appear which are already in the tables. These formulas are therefore more convenient than S' and B'', in which mean values of two successive derivatives appear. If we want to use these formulas, it is advisable to tabulate these mean values. These are called *intermediate values*, in contrast to the *principal values* already in the table. The mean values are tabulated in the difference scheme before beginning differentiation. This has been done in italics in the example given in the next paragraph.

5. *Example*: In the digging of a mine shaft, for each drill, the square of the velocity is plotted as a function of the time by means of a Karlik velocity meter. From such a diagram we record the velocity second by second for the beginning of a drill. The acceleration is to be calculated from this table.

t sec	v m/sec	Δ^1	Δ^2	Δ^3	$B'(x)$	$S'(x)$	$\int_0^1 B(t)dt$	$\int_{-1}^{+1} S(t)dt$
0	0							
	1.275	2.55	*−0.24*	−0.02			1.295	
1	2.55	*2.475*	−0.25					5.017
	3.700	2.30	*−0.275*	−0.03	2.301		3.723	
2	4.85	*2.160*	−0.28	*−0.035*		*2.166*		
	5.860	2.02	*−0.300*	−0.04	2.022		*5.885*	
3	6.87	*1.860*	−0.32	*−0.020*		*1.863*		13.633
	7.720	1.70	*−0.320*	−0.00	1.700		*7.747*	
4	8.57	*1.540*	−0.32	*+0.010*		*1.538*		
	9.260	1.38	*−0.310*	+0.02	1.379		*9.286*	
5	9.95	*1.230*	−0.30	*+0.035*		*1.224*		19.800
	10.490	1.08	*−0.275*	+0.05	1.078		*10.513*	
6	11.03	*0.955*	−0.25	*+0.070*		*0.943*		
	11.445	0.83	*−0.195*	+0.09	0.826		*11.461*	
7	11.86	*0.760*	−0.14	*+0.095*		*0.744*		23.673
	12.255	0.69	*−0.090*	+0.10	0.686		*12.263*	
8	12.65	*0.670*	−0.04	*+0.095*		*0.654*		
	12.975	0.65	*+0.005*	+0.09	0.646		*12.975*	

t sec	v m/sec	Δ^1	Δ^2	Δ^3	$B'(x)$	$S'(x)$	$\int_0^1 B(t)dt$	$\int_{-1}^{+1} S(t)dt$
9	13.30	*0.675*	+0.05	*+0.050*		*0.667*		26.617
	13.650	0.70	*+0.045*	+0.01	0.700		*13.646*	
10	14.00	*0.720*	+0.04	*−0.080*		*0.733*		
	14.370	0.74	*−0.045*	−0.17	0.747		*14.374*	
11	14.74	*0.675*	−0.13	*−0.180*		*0.705*		29.437
	15.045	0.61	*−0.225*	−0.19	0.618		*15.064*	
12	15.35	*0.450*	−0.32	*−0.085*		*0.464*		
	15.495	0.29	*−0.310*	+0.02	0.289		*15.521*	
13	15.64	*0.140*	−0.30	*+0.140*		*0.117*		31.100
	15.635	−0.01	*−0.170*	+0.26	−0.021		*15.649*	
14	15.63	*−0.030*	−0.04	*+0.160*		*−0.053*		
	15.605	−0.05	*−0.010*	+0.06	−0.053		*15.606*	
15	15.58	*−0.040*	+0.02					31.166
	15.565	−0.03	*−0.055*	−0.15			*15.570*	
16	15.55						*180.578* m	180.524 m.

The acceleration values for the half second, which are calculated from the principle values, are given in ordinary print, while the acceleration values for the seconds, which are calculated from Stirling's formula, are given in italics. If the accelerations are desired for the beginning and end of an interval, the scheme must be supplemented by a standard third difference.

6. With the first two formulas (8) and (9), a problem may be considered which resembles that of 10.15. Many measuring devices do not give the value of the function to be measured, but a mean value of the desired function over finite time intervals. For example, many anemometers give only hourly values, while instantaneous values can be read from the other meteorological instruments. For comparison, *instantaneous values* must be derived from the *hourly values*. The given mean values of the function $f(x)$ are $1/h \int_{x_0}^{x_0+h} f(x)\,dx$. We now set

$$F(x+h) - F(x) = \int_{x_0}^{x_0+h} f(x)\,dx = \Delta^1_{1/2}F;$$

(10)

$$\int_{x_0+h}^{x_0+2h} f(x)\,dx = \Delta^1_{3/2}F \cdots .$$

The difference table is then constructed from these first differences, which are taken from the listed values. Then the integral function $F(x)$ can be approximated by the Bessel formula or by the Stirling formula. From this, the derivative $F'(x)$, i.e., the instantaneous value of $f(x)$ can be obtained most simply for the center or ends of the interval from (8) and (9).

7. In numerical differentiation, as well as in integration, we are generally not concerned with the convergence of the interpolation series. We care only that the interval width is so chosen that a difference of not too high an order will be small enough to be neglected. It can readily be seen how doubtful is this calculation for differentiation because of the appearance of the interval width h or h^2 in the denominator. A good approximation of a function is not necessarily a good approximation of its derivatives, as is easily made clear by a figure. If we have an absolutely convergent series, in which termwise differentiation is permitted, then on one hand we must not take the interval width h too small. Otherwise, for the attainment of sufficient accuracy, we would be required to know the denominator h or h^2 to too many places for the individual function values. On the other hand, the intervals must not be made too large, since then we must use too many differences, some of which cannot be determined as a consequence of the propagation of the rounding off error. Then in each case there is an optimum value of the interval length, which is not generally specified. If we are dealing with divergent series, then the above formulas can still be used for the numerical calculations. In this case the differences will fall off very rapidly at first to a very small value. Then they will increase again. The series up to the smallest difference can give a useful approximation if the high differences (which increase in value) are neglected. We must therefore see to it that we obtain a useful approximation with the smallest possible number of terms. We therefore disregard the possibility of obtaining general expressions for the coefficients of the differentiation formulas, as is possible for the case of equidistant argument values (by means of Bernoulli polynomials). Instead we must content ourselves with the first members of the series given above.[3]

NOTES

1. For example, Steffensen, *Interpolationslaere* (Copenhagen 1925), Art. 7.
2. For example, Lindow, *Numerische Infinitesimalrechnung* (Be[illegible] p. 166 ff.
3. Nörlund, *Differenzenrechnung* (Berlin, 1924), Ch. VIII, Art. 7.

12. Numerical Integration.

1. Each of the interpolation series derived above can be used for the *numerical integration* of a function given in tables. We do not consider the development of formulas for tables with unequally distant function values, since they are not clearly arranged, and the calculation required in their use is very intricate. In comparison, the integration of a function for which a difference scheme of equidistant values already exists requires

very little work and yields an estimate of the error resulting in the integration.

2. The integration formula obtained from the Newton interpolation formula N_+ had already been given by Gregory in a letter of Nov. 23, 1670 to Collins. It is of little interest to us and hence will not be derived here.

On the other hand, in the integration of differential equations, 32.7, the integration series resulting from Newton's formula N_- is often used. If this formula is integrated termwise over an interval, the coefficient of the mth difference becomes

$$A_m = \int_0^1 \binom{t + m - 1}{m} dt.$$

If this is evaluated, we obtain[1]

$$\int_{x_0}^{x_0+h} f(x)\,dx = h\int_0^1 N_-(\tau)\,dt = h\Big[y_0 + \frac{1}{2}\Delta^1_{-1/2} + \frac{5}{12}\Delta^2_{-1}$$

$$(1) \qquad + \frac{3}{8}\Delta^3_{-3/2} + \frac{251}{720}\Delta^4_{-2} + \frac{95}{288}\Delta^5_{-5/2} + \frac{19087}{60480}\Delta^6_{-3}$$

$$+ \frac{5257}{17280}\Delta^7_{-7/2} + \cdots\Big] + R_n\,, \quad \text{where } -n \leqq \lambda t \leqq +1.$$

$$(2) \qquad R_n = h^{n+1}\int_0^1 \frac{t(t+1)\cdots(t+n-1)}{(n)!} f^{(n)}(\lambda t)\,dt.$$

If $f^{(n)}(\lambda t)$ is continuous, the mean value theorem of the integral calculus can be used, since the other factors of the integrand do not change their signs from 0 to 1. We then obtain a somewhat different form of the remainder term,

$$(3) \qquad R_n = h^{n+1}f^{(n)}(\tau)\int_0^1 \binom{t+n-1}{n} dt = Ch^{n+1}f^{(n)}(\tau),$$

the function $f($ [illegible]

[illegible] $-n \leqq \tau \leqq +1.$

Integration can be performed over two intervals, from $x_0 - h$ to $x_0 + h$, just as well as over one interval. We then get a formula which is especially distinguished by convenient coefficients for the differences. This formula is

$$\int_{x_0-h}^{x_0+h} f(x)\,dx = h\int_{-1}^{+1} N_-(t)\,dt = 2hy_0 + h\Big[\frac{1}{3}\Delta^2_{-1} + \frac{1}{3}\Delta^3_{-3/2}$$

$$(4) \qquad + \frac{29}{90}\Delta^4_{-2} + \frac{14}{45}\Delta^5_{-5/2} + \frac{1139}{3780}\Delta^6_{-3} + \frac{123}{420}\Delta^7_{-7/2}\cdots\Big] + R_n\,,$$

where, for the application of the mean value theorem the remainder term

is to be divided into two parts. For practical calculations, the above formula is best used in the form

$$\int_{x_0-h}^{x_0+h} f(x)\,dx = 2hy_0 + \frac{1}{3}h\Big[\Delta^2_{-1} + \Delta^3_{-3/2} + \Delta^4_{-2} + \Delta^5_{-5/2} + \Delta^6_{-3} \tag{5}$$

$$+ \Delta^7_{-7/2} + \cdots - \frac{1}{30}\Delta^4_{-2} - \frac{1}{15}\Delta^5_{-5/2} - \frac{121}{1260}\Delta^6_{-3}$$

$$- \frac{17}{140}\Delta^7_{-7/2}\cdots\Big] + R_n .$$

3. For ordinary quadrature, it is customary to use the formulas which follow from Bessel's and Stirling's formulas. These require less work of calculation because the coefficients of the high powers diminish rapidly.

The integration of the Bessel formula, 10(28), over an interval gives

$$\int_{x_0}^{x_0+h} f(x)\,dx = h\int_0^1 B(t)\,dt = h\Big(y_0 + \frac{1}{2}\Delta^1_{1/2} - \frac{1}{12}\overline{\Delta^2_{1/2}} \tag{6}$$

$$+ \frac{11}{720}\overline{\Delta^4_{1/2}} - \frac{191}{60480}\overline{\Delta^6_{1/2}} + \frac{2497}{3628800}\overline{\Delta^8_{1/2}}\cdots\Big) + R_{2n} ,$$

where here, for $-n + 1 \leqq \tau \leqq n$

$$R_{2n} = h^{2n+1}Cf^{(2n)}(\tau), \tag{7}$$

if C denotes the integral over the corresponding function of the Bessel interpolation formula. The use of the mean value theorem is possible, since this function does not change its sign in the interval of integration.

If we want to calculate the value of the integral for a larger interval, and if the limits of this interval coincide with the limits of the entire interval, then we can write down the above formula for a series of successive intervals and sum these terms. We then get

$$\int_{x_0}^{x_0+\kappa h} f(x)\,dx = h\Big\{\sum_{\lambda=0}^{\kappa-1} y_\lambda + \frac{1}{2}\sum_{\lambda=0}^{\kappa-1}\Delta^1_{\lambda+1/2} - \frac{1}{12}\sum_{\lambda=0}^{\kappa-1}\Delta^2_{\lambda+1/2} \tag{8}$$

$$+ \frac{11}{720}\sum_{\lambda=0}^{\kappa-1}\overline{\Delta^4_{\lambda+1/2}}\cdots\Big\} + R_{2n} .$$

Now, since $\Delta^1_{\lambda+1/2} = y_{\lambda+1} - y_\lambda$, then $\sum_{\lambda=0}^{\kappa-1}\Delta_{\lambda+1/2} = y_\kappa - y_0$ so that the first two terms together give $\sum_{\lambda=0}^{\kappa} y_\lambda - (y_0 + y_\kappa)/2$. The other sums may also be expressed by the differences of differences of lower order. For example,

$$\sum_{\lambda=0}^{\kappa-1}\Delta^2_\lambda = \Delta^1_{\kappa-1/2} - \Delta^1_{-1/2} \quad \text{and} \quad \sum_{\lambda=0}^{\kappa-1}\Delta^2_{\lambda+1} = \Delta^1_{\kappa+1/2} - \Delta^1_{1/2} \tag{8a}$$

etc. If we introduce this in the above formula, we have

$$
\begin{aligned}
\int_{x_0}^{x_0+\kappa h} f(x)\,dx = h\Bigg\{ \sum_{\lambda=0}^{\kappa} y_\lambda - \frac{y_0+y_\kappa}{2} - \frac{1}{12}(\overline{\Delta_\kappa^1} - \overline{\Delta_0^1}) \\
+ \frac{11}{720}(\overline{\Delta_\kappa^3} - \overline{\Delta_0^3}) \cdots \Bigg\} - \frac{191}{60480}(\overline{\Delta_\kappa^5} - \overline{\Delta_0^5}) \qquad (9) \\
+ \frac{2497}{3628800}(\overline{\Delta_\kappa^7} - \overline{\Delta_0^7}) + \cdots + R_{2n},
\end{aligned}
$$

where $\overline{\Delta_\kappa^m} = 1/2(\Delta_{\kappa+1/2}^m + \Delta_{\kappa-1/2}^m)$, and where, for $-(n-1) \leqq \bar{\tau} \leqq (n+\kappa)$,

$$R_{2n} = \kappa h^{2n+1} C f^{(2n)}(\bar{\tau}).$$

4. *Example:*[3] In the example of the compressed spring given in 10.4, the energy stored in the spring is to be computed as a function of the contraction. Since the intermediate values are found in the Bessel form, these are introduced first. This gives the table below, in which the intermediate values are given in italics:

TABLE I.

h	y	Δ^1	Δ^2	Δ^3	[ ]	E mmkg	E mmkg
		44					
0	0	*46.35*	4.7				
	24.35	48.7	*6.25*	3.1	*23.83*	*119.15*	
5	48.7	*52.6*	7.8				500
	76.95	56.5	*9.25*	2.9	*76.18*	*380.90*	
10	105.2	*61.85*	10.7				
	138.80	67.2	*12.25*	3.1	*137.78*	*688.90*	
15	172.4	*74.10*	13.8				1747
	212.90	81	*15.50*	3.4	*211.61*	*1058.05*	
20	253.4	*89.60*	17.2				
	302.50	98.2	*18.55*	2.7	*300.95*	*1504.75*	
25	351.6	*108.15*	19.9				3549
	410.65	118.1	*21.50*	3.2	*408.86*	*2044.30*	
30	469.7	*129.65*	23.1				
	540.30	141.2	*24.75*	3.3	*538.24*	*2691.20*	
35	610.9	*154.40*	26.4				6153
	694.70	167.6	*27.95*	3.1	*692.37*	*3461.85*	
40	778.5	*182.35*	29.5				
		197.1				11949.10	
Σ	2790.4				2389.82		11949

To calculate the integral for the first interval, we must interpolate above the lines drawn in the table which limit the original scheme. The intermediate values here are also formed. The column designated

by [ ] gives the value of the integral, not yet multiplied by h, of the individual intervals, while the next column gives the actual value of the integral in mm. kg. The total stored energy has the value of 11.95 m. kg. As a check on the calculation, we make use of the Bessel formula for n intervals. This gives the following:

$$
\begin{aligned}
\sum_0^8 y &= 2790.4 \\
-\frac{1}{2}(y_0 + y_8) &= -389.25 \\
-\frac{1}{24}(\Delta_{15/2} + \Delta_{17/2}) &= -15.20 \\
+\frac{1}{24}(\Delta_{-1/2} + \Delta_{1/2}) &= 3.86 \\
\hline
[\quad] &= 2389.81.
\end{aligned}
$$

This is in good agreement with the sum of the corresponding column of the table.

As a second example, we consider the completely analogous table of changes in velocity of 11.5. In the column labeled $\int_0^1 B(t)\,dt$ are tabulated the distances covered in each second. As a check on the sum 180.578 m., we again use the formula for n intervals. In this case

$$\int_0^{16} B(t)\,dt = 180.56.$$

Since this agrees within the limit of error, the total values are therefore checked.

If the limits of the integration interval do not coincide with the limits of the interval of the difference scheme, we must use tables for the functions which appear as factors of the differences, or we must calculate the value of the integral up to the values of the argument which are adjacent to the limits of the integral. We then form a difference scheme, and interpolate in this scheme to the limits of integration.

5. We have previously mentioned that *Stirling's formula*, 10(23), is most profitably used for integration over a double interval. This gives

$$
\begin{aligned}
\int_{x_0-h}^{x_0+h} f(x)\,dx = h\int_{-1}^{+1} f(t)\,dt = 2h\Big[y_0 + \frac{\Delta_0^2}{6} - \frac{\Delta_0^4}{180} + \frac{\Delta_0^6}{1512} \\
- \frac{23\Delta_0^8}{226800}\cdots\Big] + R_{2n}. \qquad (-n \leqq \tau \leqq +n)
\end{aligned}
\tag{10}
$$

This formula is simpler than that of Bessel, as it contains only principal values. The remainder term again has the form

$$R_{2n} = h^{2n+1} C f^{(2n)}(\tau), \tag{11}$$

by application of the mean value theorem. Here C is the integral extended over the corresponding term of Stirling's formula.

The values (in the last column) for the energy of the spring in the preceding section and for the distance in 11.5 are calculated by means of this formula. The sum of the energy values calculated in this way is in satisfactory agreement with that calculated by the Bessel formula. If we terminate Stirling's integral formula with differences of the second order, we obtain, starting from y_1,

$$\begin{aligned} \int_{x_1-h}^{x_1+h} f(x)\,dx &= 2h\left(y_1 + \frac{\Delta_1^2}{6}\right) = 2h\left(y_1 + \frac{1}{2}(y_0 - 2y_1 + y_2)\right) \\ &= \frac{h}{3}(y_0 + 4y_1 + y_2). \end{aligned} \tag{12}$$

This is the so-called *Kepler barrel rule*, used by him in computing the volume of a barrel. If we sum this formula for several double intervals, we obtain Simpson's rule (cf. 15.4):

$$\begin{aligned} y &= \frac{h}{3}[y_0 + 4y_1 + 2y_2 + 4y_3 + 2y_4 \cdots] \\ &= \frac{h}{3}\left[4 \sum_0^{n-1} y_{2k+1} + 2 \sum_0^{n} y_{2k} - (y_0 + y_{2n})\right]. \end{aligned} \tag{13}$$

Stirling's integral formula is then an improved Simpson's rule.

6. If we want to estimate the errors arising in integration, we first form the corresponding derivative of the function. From this we obtain the maximum value of the function in the given region and substitute this in the formula for the remainder term. This will give an upper limit to the error. But in general these derivatives are difficult to calculate. It is therefore customary to take the difference corresponding to 10(5) as an approximate value of the derivative, since $\Delta^l_{m+l/2} = h^l f^{(l)}(\xi)$, and to substitute the maximum value of this, in the interval considered, in the remainder term.

If we are dealing with a rational integral function of nth degree, and if we take the interpolation formulas up to the nth difference, all the formulas derived above must give exact results, aside from some rounding off errors, since the remainder term is zero.

If this is not the case, then the error can still be made arbitrarily small by a corresponding reduction of the argument interval, in the case of integration over a fixed interval from $a = x_0$ to $b = x_0 + \kappa h$. The reason for this is that the error term has the factor h^{n+1} or h^{2n+1} for each subinterval. Therefore, if the maximum value of the other factors is denoted by M, the error for the entire interval a to b is smaller than $Mh^n\,(b - a)$ or $Mh^{2n}(b - a)$. Each of these becomes arbitrarily small with h if the corresponding derivative is finite.

In order to compare, in a particular case, the error estimated by means of this remainder term with the actual error which occurs, we form the difference scheme for $10^7 \cos x$ which is not reproduced here. The integral of $\cos x\, dx$ is $\sin x +$ constant, so that the values of the integral can easily be checked. From the table it can immediately be seen that there is no point in going beyond the 8th difference, since the error due to rounding off would be considerable there (cf. 10.5).

Next, let us take the Bessel formula for the entire interval from 0 to π, so that we first introduce the intermediate values and take 80 as the maximum value of Δ^8 from $-36°$ to $+216°$, provided we still consider the fifth differences in the calculation. Then

$$R = \frac{2497}{3628800}\pi\left(\frac{\pi}{15}\right)^8 f^{(8)}(\tau) = \frac{2497}{3628800}\pi\Delta^8 \leqq \frac{2497 \times 80\pi}{3628800} \approx 0.2.$$

The error therefore remains less than 1/5 of a unit in the last place. If we carry out the calculation for any part of the interval, we find in fact the exact value, just as we get exact values if we calculate the values of the function for the individual intervals. That an exact value is obtained for the π interval is understandable because of the symmetry.

We now use Stirling's formula and calculate the value of the integral for the individual double intervals. If we use the differences up to the fourth, then the error of the sum of these integrals is

$$R \leqq \frac{\pi}{2}\frac{1}{1512} \times 828 \approx 0.8.$$

This is an error of about one in the last place. The error in fact is about one in the last place since the error of rounding off also enters in. This error fluctuates, being sometimes positive and sometimes negative.

The error is considerably larger in the two formulas originally derived from the interpolation series N_-. The values calculated by both formulas and the corresponding errors are reproduced in the

accompanying table. The numbers in italics are the values of the integral for the single or double integral; the other numbers are the values of the integral from zero on, which have been obtained by addition.

TABLE II.

φ	$10^7 \sin x$	Newton for 1 interval	Error	Newton for 2 Intervals	Error
0	0000000	0000000	0	0000000	0
12	2079116	2079121	+ 5		
		1988255			
24	4067366	4067376	+10	4067371	+ 5
		1810486			
36	5877852	5877862	+10		
		1553596		*3364081*	
48	7431449	7431458	+ 9	7431452	+3
		1228801			
60	8660253	8660259	+ 6		
		850307		*2079113*	
72	9510565	9510566	+ 1	9510565	0
		434645			
84	9945217	9945211	− 6		
		− *8*		*0434646*	
96	9945217	9945203	−14	9945211	− 6
		− *434664*			
108	9510565	9510539	−26		
		− *850321*		−*1284974*	
120	8660253	8660218	−35	8660237	−16
		−*1228818*			
132	7431449	7431400	−49		
		−*1553605*		−*2782406*	
144	5877852	5877795	−59	5879831	−21
		−*1810497*			
156	4067366	4067298	−68		
		−*1988257*		−*3798744*	
168	2079116	2079041	−75	2079087	−29
		−*2079126*			
180	0000000	− 85	−85		

An upper limit for the error may be given just as before. In this calculation, the sixth difference is still employed. Since the maximum value of Δ^7 is about 200, the maximum error for the entire interval, using single intervals with the Newton formula, is

$$R \leqq \pi \times \frac{5257}{17280} \times 200 \approx 190,$$

and using the Newton integration formula for double intervals,

$$R \leqq \pi \times \frac{123}{420} \times 200 \approx 185.$$

The errors then are within the maximum errors previously found, as was to be expected. The second of the last two formulas is considerably easier to handle, and consequently the results are obtained more rapidly. For this reason, since the errors to be expected are about equal in the two cases, the second method is to be preferred to the first. Its only disadvantage is that the interval between values of the function calculated is twice as large as that between the given argument values.

7. We can get approximate values for multiple integrals in the same way as for single integrals, i.e., by a multiple integration of the interpolation formula. For example, if the double integral is denoted by $\overline{\overline{J}}$, and the single integral by $\overline{J}$, we have

$$\overline{\overline{J}} = \overline{\overline{J}}_0 + \overline{J}_0(x - x_0) + \int_{x_0}^{x} \int_{x_0}^{x} f(x)\, dx^2. \tag{14}$$

If the new variable t is introduced, and the integration is performed over a single interval, using the Newton formula N_-, we have

$$\begin{aligned} \overline{\overline{J}}_1 &= \overline{\overline{J}}_0 + \overline{J}_0 h + h^2 \int_0^1 \int_0^t \binom{t + i - 1}{i} \Delta^i_{-i/2}\, dt^2 \\ &= \overline{\overline{J}}_0 + \overline{J}_0 h + h^2 \left[\frac{y_0}{2} + \frac{1}{6} \Delta^1_{-1/2} + \frac{1}{8} \Delta^2_{-1} + \frac{19}{180} \Delta^3_{-3/2} \right. \\ &\quad + \frac{3}{32} \Delta^4_{-2} + \frac{863}{10080} \Delta^5_{-5/2} + \frac{275}{3456} \Delta^6_{-3} \\ &\quad \left. + \frac{33953}{453600} \Delta^7_{-7/2} \cdots \right] + R_n . \end{aligned} \tag{15}$$

After one value has been obtained with this formula, further calculations can be carried out more conveniently with the following formula, developed by Adams and Störmer.[4] If the above integration is performed over the interval from 0 to $-h$, then

$$\begin{aligned} \overline{\overline{J}}_{-1} &= \overline{\overline{J}}_0 - \overline{J}_0 h + h^2 \left[\frac{y_0}{2} - \frac{1}{6} \Delta^1_{-1/2} - \frac{1}{24} \Delta^2_{-1} - \frac{2}{90} \Delta^3_{-3/2} - \frac{7}{480} \Delta^4_{-2} \right. \\ &\quad \left. - \frac{107}{10080} \Delta^5_{-5/2} - \frac{199}{24192} \Delta^6_{-3} - \frac{6031}{907200} \Delta^7_{-7/2} \cdots \right] + \overline{R}_n . \end{aligned} \tag{16}$$

If these two equations are added together, a formula is obtained from which $\bar{J}_0$ is absent and in which the two terms containing the double integral remain:

$$\begin{aligned}\bar{\bar{J}}_1 = 2\bar{\bar{J}}_0 - \bar{\bar{J}}_{-1} + h^2\Big[y_0 + \frac{1}{12}\Delta^2_{-1} + \frac{1}{12}\Delta^3_{-3/2} + \frac{19}{240}\Delta^4_{-2} + \frac{3}{40}\Delta^5_{-5/2}\\ + \frac{863}{12096}\Delta^6_{-3} + \frac{275}{4032}\Delta^7_{-7/2}\cdots\Big] + R_n + \bar{R}_n .\end{aligned} \tag{17}$$

For practical calculation, it is better to write this formula in the form

$$\begin{aligned}\bar{\bar{J}}_1 = 2\bar{\bar{J}}_0 - \bar{\bar{J}}_{-1} + h^2 y_0 + \frac{h^2}{12}\Big[\Delta^2_{-1} + \Delta^3_{-3/2} + \Delta^4_{-2} + \Delta^5_{-5/2} + \Delta^6_{-3}\\ + \Delta^7_{-7/2} + \cdots - \frac{1}{20}\Delta^4_{-2} - \frac{1}{10}\Delta^5_{-5/2} - \frac{145}{1008}\Delta^6_{-3}\\ - \frac{61}{336}\Delta^7_{-7/2}\cdots\Big] + R_n + \bar{R}_n .\end{aligned} \tag{18}$$

Expressions are obtained for R_n and $\bar{R}_n$ which are completely analogous to those obtained for single integrals:

$$R_n + \bar{R}_n = h^{r+2}(Af^{(n)}(\tau) + Bf^{(n)}(\bar{\tau})), \tag{19}$$

where $-(n-1) \leqq \tau \leqq 1$; $-n \leqq \tau \leqq 0$, and where A and B are double integrals of the corresponding polynomial of the interpolation series.

Formulas for double integrals, employing other interpolation formulas, can be constructed in exactly the same way. If, for example, the Stirling formula is first integrated from 0 to 1 and then from 0 to -1, there results

$$\begin{aligned}\bar{\bar{J}}_1 = \bar{\bar{J}}_0 + \bar{J}_0 h + h^2\left(\frac{y_0}{2} + \frac{\overline{\Delta^1_0}}{6} + \frac{\Delta^2_0}{24} - \frac{7\overline{\Delta^3_0}}{360} - \frac{\Delta^4_0}{480} + \cdots\right) + R_{2n} ,\\ \bar{\bar{J}}_{-1} = \bar{\bar{J}}_0 + \bar{J}_0 h + h^2\left(\frac{y_0}{2} - \frac{\overline{\Delta^1_0}}{6} + \frac{\Delta^2_0}{24} + \frac{7\overline{\Delta^3_0}}{360} - \frac{\Delta^4_0}{480} + \cdots\right) + \bar{R}_{2n} .\end{aligned} \tag{19a}$$

By addition of these formulas, we obtain a formula due to Legendre:[5]

$$\begin{aligned}\bar{\bar{J}}_1 = 2\bar{\bar{J}}_0 - \bar{\bar{J}}_{-1} + h^2\left(y_0 + \frac{\Delta^2_0}{12} - \frac{\Delta^4_0}{240} + \frac{31\Delta^6_0}{60480} - \frac{289\Delta^8_0}{3628800}\cdots\right)\\ + R_{2n} + \bar{R}_{2n} .\end{aligned} \tag{20}$$

The remainder term has the form

$$R_{2n} + \overline{R}_{2n} = 2h^{2n+2}\cdot f^{(2n)}(\tau)$$

(21)

$$\cdot \int_0^1 \int_0^t \frac{t^2(t^2 - 1^2)(t^2 - 2^2)\cdots(t^2 + (n-1)^2)}{(2n)!}\,dt^2,$$

where $-n \leqq \tau \leqq +n$. We omit construction of additional formulas here. There is a series of such formulas, corresponding to the interpolation formulas, the region of integration of which can be extended to a larger region by summation.

Example: If the formula derived from N_- for the single interval is applied to $10^7 \cos x$, we obtain

$$\int_0^{\pi/15} \int_0^x 10^7 \cos x\,dx = -10^7 + 218{,}524 = 9{,}781{,}476,$$

which is also the exact value of $-10^7 \cos \pi/15$, if the differences are employed as far as the sixth. From there on, the formula of Adams may well be used. This gives, for example, with four differences,

$$-10^7 \cos\frac{2\pi}{15} = \int_0^{2\pi/15} \int_0^x 10^7 \cos x\,dx$$

$$= -19{,}562{,}952 + 10{,}000{,}000 + 427{,}494$$

$$= -9{,}135{,}458,$$

while the exact value is 9,135,454. If six differences are used, the exact value is again obtained. These calculations should not be carried out any further, any more than in the estimations of the errors, since they can contribute nothing new.

NOTES

1. Bashforth and Adams, *An Attempt to Test the Theories of Capillary Action* (Cambridge, 1883).
2. Cf. note 2 in Art. 10.
3. Lindow, *Numerische Infinitesimalrechnung* (Berlin and Bonn, 1928), p. 173 ff.
4. Störmer, *Archives des sciences physiques et naturelles* (Geneva, 1907), July to October, p. 63.
5. Legendre, *Traité des fonctions élliptiques* II (Paris, 1826), p. 52.

13. Interpolation with Functions of Several Variables. Cubature.

1. The interpolation with functions of several variables is extremely involved for arbitrarily chosen values of the argument. This results from the fact that it is scarcely possible to prepare tables with any degree of clarity for completely arbitrary values of the argument. With functions

of two variables, the graphical representation, as is used in weather maps, is the most clearly arranged. But even with this mode of representation, the clarity is greatly reduced with three variables, and for more than three, a representation of this type is worthless.

If data are to be prepared from observations, it is desirable, and frequently possible, so to arrange the observations that all the variables are kept constant but one, and that this one is always changed by a constant increment of its argument. The magnitude of the increment for different series of observations is chosen as nearly the same as possible. We then obtain, for two variables, the *lattice points* of a rectangular array, as is

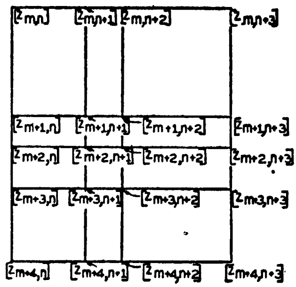


FIG. 33

shown in Fig. 33. We limit ourselves here to this case, and to the case of two variables. How such formulas will appear for more variables can be seen from the formulas derived here. The *function* and *difference symbols* will be the same as in the preceding articles, but the position of any quantity must now be characterized by two indices, as is shown for the values of the function in Fig. 33.

2. First we construct a *difference scheme* for the function of two variables, corresponding to the difference scheme of 9.1. Here we use the symbol δ for the divided differences, to be distinguished from Δ, which is used for ordinary differences for equidistant values. We then form the divided difference in the direction of both variables, x and y, and choose the two superscripts so that the first index gives the order of the divided difference for x, the second gives the order for y, while the subscript, as has been mentioned above, denotes the position in the difference scheme. It therefore reveals the values from which the differences are formed. The first differences become

$$(1) \qquad \delta^{10}_{m+1/2,n} = \frac{z_{m,n} - z_{m+1,n}}{x_m - x_{m+1}}; \qquad \delta^{01}_{m,n+1/2} = \frac{z_{m,n} - z_{m,n+1}}{y_n - y_{n+1}}.$$

From these two first differences we can form the three second differences:

$$
\begin{aligned}
\delta^{20}_{m+1,n} &= \frac{\delta^{10}_{m+1/2,n} - \delta^{10}_{m+3/2,n}}{x_m - x_{m+2}} = \frac{z_{m,n}}{(x_m - x_{m+1})(x_m - x_{m+2})} \\
&\quad + \frac{z_{m+1,n}}{(x_{m+1} - x_m)(x_{m+1} - x_{m+2})} + \frac{z_{m+2,n}}{(x_{m+2} - x_m)(x_{m+2} - x_{m+1})}, \\
\delta^{11}_{m+1/2,n+1/2} &= \frac{\delta^{10}_{m+1/2,n} - \delta^{10}_{m+1/2,n+1}}{y_n - y_{n+1}} = \frac{\delta^{01}_{m,n+1/2} - \delta^{01}_{m+1,n+1/2}}{x_m - x_{m+1}} \\
&= \frac{z_{m,n}}{(x_m - x_{m+1})(y_n - y_{n+1})} + \frac{z_{m+1,n}}{(x_{m+1} - x_m)(y_n - y_{n+1})} \\
&\quad + \frac{z_{m,n+1}}{(x_m - x_{m+1})(y_{n+1} - y_n)} + \frac{z_{m+1,n+1}}{(x_{m+1} - x_m)(y_{n+1} - y_n)}, \\
\delta^{02}_{m,n+1} &= \frac{\delta^{01}_{m,n+1/2} - \delta^{01}_{m,n+3/2}}{y_n - y_{n+2}} = \frac{z_{m,n}}{(y_n - y_{n+1})(y_n - y_{n+2})} \\
&\quad + \frac{z_{m,n+1}}{(y_{n+1} - y_n)(y_{n+1} - y_{n+2})} + \frac{z_{m,n+2}}{(y_{n+2} - y_n)(y_{n+2} - y_{n+1})}.
\end{aligned}
\tag{2}
$$

From the latter expressions, it follows that the order of the terms, in which the values of the functions appear in the differences, is arbitrary. This theorem can be proved, in general, in a way similar to the corresponding theorem for one variable (8.2). We also write down the formulas for the differences of third order:

$$
\begin{aligned}
\delta^{30}_{m+3/2,n} &= \frac{\delta^{20}_{m+1,n} - \delta^{20}_{m+2,n}}{x_m - x_{m+3}}, \\
\delta^{21}_{m+1,n+1/2} &= \frac{\delta^{11}_{m+1/2,n+1/2} - \delta^{11}_{m+3/2,n+1/2}}{x_m - x_{m+2}} = \frac{\delta^{20}_{m+1,n} - \delta^{20}_{m+1,n+1}}{y_n - y_{n+1}}, \\
\delta^{12}_{m+1/2,n+1} &= \frac{\delta^{11}_{m+1/2,n+1/2} - \delta^{11}_{m+1/2,n+3/2}}{y_n - y_{n+2}} = \frac{\delta^{02}_{m,n+1} - \delta^{02}_{m+1,n+1}}{x_m - x_{m+1}}, \\
\delta^{03}_{m,n+3/2} &= \frac{\delta^{02}_{m,n+1} - \delta^{02}_{m,n+2}}{y_n - y_{n+3}}, \text{ etc.}
\end{aligned}
\tag{3}
$$

If these divided differences are formed, we obtain the following scheme:

	0	$\frac{1}{2}$	1	$\frac{3}{2}$	2
0	$[z_{00}]$	$\delta^{01}_{0(1/2)}$	$[z_{01}]$ δ^{02}_{01}	$\delta^{01}_{0(3/2)}$ $\delta^{03}_{0(3/2)}$	$[z_{02}]$ δ^{02}_{02} δ^{04}_{02}
$\frac{1}{2}$	$\delta^{10}_{(1/2)0}$	$\delta^{11}_{(1/2)(1/2)}$	$\delta^{10}_{(1/2)1}$ $\delta^{12}_{(1/2)1}$	$\delta^{11}_{(1/2)(3/2)}$ $\delta^{13}_{(1/2)(3/2)}$	$\delta^{10}_{(1/2)2}$
1	$[z_{10}]$ δ^{20}_{10}	$\delta^{01}_{1(1/2)}$ $\delta^{21}_{1(1/2)}$	$[z_{11}]$ δ^{02}_{11} δ^{20}_{11} δ^{22}_{11}	$\delta^{01}_{1(3/2)}$ $\delta^{03}_{1(3/2)}$ $\delta^{21}_{1(3/2)}$	$[z_{12}]$ δ^{02}_{12} δ^{04}_{12} δ^{20}_{12} δ^{22}_{12}
$\frac{3}{2}$	$\delta^{10}_{(3/2)0}$ $\delta^{30}_{(3/2)0}$	$\delta^{11}_{(3/2)(1/2)}$ $\delta^{31}_{(3/2)(1/2)}$	$\delta^{10}_{(3/2)1}$ $\delta^{12}_{(3/2)1}$ $\delta^{30}_{(3/2)1}$	$\delta^{11}_{(3/2)(3/2)}$ $\delta^{13}_{(3/2)(3/2)}$ $\delta^{31}_{(3/2)(3/2)}$	$\delta^{10}_{(3/2)2}$ $\delta^{12}_{(3/2)2}$ $\delta^{30}_{(3/2)2}$
2	$[z_{20}]$ δ^{20}_{20} δ^{40}_{20}	$\delta^{01}_{2(1/2)}$ $\delta^{21}_{2(1/2)}$	$[z_{21}]$ δ^{02}_{21} δ^{20}_{21} δ^{22}_{21} δ^{40}_{21}	$\delta^{01}_{2(3/2)}$ $\delta^{03}_{2(3/2)}$ $\delta^{21}_{2(3/2)}$	$[z_{22}]$ δ^{02}_{22} δ^{04}_{22} δ^{20}_{22} δ^{22}_{22} δ^{40}_{22}
$\frac{5}{2}$	$\delta^{10}_{(5/2)0}$ $\delta^{30}_{(5/2)0}$	$\delta^{11}_{(5/2)(1/2)}$ $\delta^{31}_{(5/2)(1/2)}$	$\delta^{10}_{(5/2)1}$ $\delta^{12}_{(5/2)1}$ $\delta^{30}_{(5/2)1}$	$\delta^{11}_{(5/2)(3/2)}$ $\delta^{13}_{(5/2)(3/2)}$ $\delta^{31}_{(5/2)(3/2)}$	$\delta^{10}_{(5/2)2}$ $\delta^{12}_{(5/2)2}$ $\delta^{30}_{(5/2)2}$
3	$[z_{30}]$ δ^{20}_{30} δ^{40}_{30}	$\delta^{01}_{3(1/2)}$ $\delta^{21}_{3(1/2)}$	$[z_{31}]$ δ^{02}_{31} δ^{20}_{31} δ^{22}_{31} δ^{40}_{31}	$\delta^{01}_{3(3/2)}$ $\delta^{03}_{3(3/2)}$ $\delta^{21}_{3(3/2)}$	$[z_{32}]$ δ^{02}_{32} δ^{04}_{32} δ^{20}_{32} δ^{22}_{32} δ^{40}_{32}

where the divided differences are given as far as the fourth. It can also be shown that these *partial divided differences* are the mean values of the corresponding partial derivatives over the interval in question, so that the general interpolation formula derived in the next section goes over into the Taylor expansion for several variables, if the coordinates of the function values are allowed to converge to a point, provided that the derivatives concerned exist and are continuous.

3. By use of the differences appearing in the scheme, a general interpolation formula can be constructed as an identity, just as in 8.3. This is

$$
(4)\qquad
\begin{aligned}
[z_{xy}] - [z_{00}] &= \frac{[z_{xy}] - [z_{0y}]}{x - x_0}(x - x_0) + \frac{[z_{0y}] - [z_{00}]}{y - y_0}(y - y_0) \\
&= \delta^{10}_{(x/2)y}(x - x_0) + \delta^{01}_{0(y/2)}(y - y_0).
\end{aligned}
$$

For the differences occurring here, we have

$$
(5)\qquad
\begin{aligned}
\delta^{10}_{(x/2)y} - \delta^{10}_{(1/2)0} &= \frac{\delta^{10}_{(x/2)y} - \delta^{10}_{(1/2)y}}{x - x_1}(x - x_1) + \frac{\delta^{10}_{(1/2)y} - \delta^{10}_{(1/2)0}}{y - y_0}(y - y_0) \\
&= \delta^{20}_{((x+1)/4)y}(x - x_1) + \delta^{11}_{(1/2)(y/2)}(y - y_0), \\
\delta^{01}_{0(y/2)} - \delta^{01}_{0(1/2)} &= \frac{\delta^{01}_{0(y/2)} - \delta^{01}_{0(1/2)}}{y - y_1}(y - y_1) = \delta^{02}_{0((y+1)/4)}(y - y_1).
\end{aligned}
$$

These differences can be further transformed. They become

$$
(6)\qquad
\begin{aligned}
\delta^{20}_{((x+1)/4)y} - \delta^{20}_{10} &= \frac{\delta^{20}_{((x+1)/4)y} - \delta^{20}_{1y}}{x - x_2}(x - x_2) + \frac{\delta^{20}_{1y} - \delta^{20}_{10}}{y - y_0}(y - y_0) \\
&= \delta^{30}_{((x+5)/8)y}(x - x_2) + \delta^{21}_{1(y/2)}(y - y_0), \\
\delta^{11}_{(1/2)(y/2)} - \delta^{11}_{(1/2)(1/2)} &= \delta^{12}_{(1/2)(y+1)/4}(y - y_1) \\
\delta^{02}_{0(y+1)/4} - \delta^{02}_{01} &= \delta^{03}_{0(y+5)/8}(y - y_2),
\end{aligned}
$$

from which we have

$$
\begin{aligned}
\delta^{30}_{((x+5)/8)y} - \delta^{30}_{(3/2)0} &= \frac{\delta^{30}_{((x+5)/8)y} - \delta^{30}_{(3/2)y}}{x - x_3}(x - x_3) \\
&\quad + \frac{\delta^{30}_{(3/2)y} - \delta^{30}_{(3/2)0}}{y - y_0}(y - y_0) \\
&= \delta^{40}_{((x+17)/16)y}(x - x_3) + \delta^{31}_{(3/2)(y/2)}(y - y_0), \\
\delta^{21}_{1(y/2)} - \delta^{21}_{1(1/2)} &= \delta^{22}_{1(y+1)/4}(y - y_1), \\
\delta^{12}_{(1/2)(y+1)/4} - \delta^{12}_{(1/2)1} &= \delta^{13}_{(1/2)(y+5)/8}(y - y_2), \\
\delta^{03}_{0(y+5)/8} - \delta^{03}_{0(3/2)} &= \delta^{04}_{0(y+17)/16}(y - y_3) \qquad \text{etc.}
\end{aligned}
$$

If the values obtained above are substituted successively in the first equation, the general interpolation formula is obtained. We write it here for the interpolation of third order with the remainder term:

$$
\begin{aligned}
[z_{x,y}] &= [z_{00}] + \delta^{10}_{(1/2)0}(x - x_0) + \delta^{01}_{0(1/2)}(y - y_0) + \delta^{20}_{10}(x - x_0)(x - x_1) \\
&\quad + \delta^{11}_{(1/2)(1/2)}(x - x_0)(y - y_0) + \delta^{02}_{01}(y - y_0)(y - y_1) \\
&\quad + \delta^{30}_{(3/2)0}(x - x_0)(x - x_1)(x - x_2) + \delta^{21}_{1(1/2)}(x - x_0)(x - x_1)(y - y_0) \\
&\quad + \delta^{12}_{(1/2)1}(x - x_0)(y - y_0)(y - y_1) \\
&\quad + \delta^{03}_{0(3/2)}(y - y_0)(y - y_1)(y - y_2) + R_4 ,
\end{aligned} \tag{8}
$$

where

$$
\begin{aligned}
R_4 = {} & \delta^{40}_{((x+17)/16)y}(x - x_0)(x - x_1)(x - x_2)(x - x_3) \\
& + \delta^{31}_{(3/2)(y/2)}(x - x_0)(x - x_1)(x - x_2)(y - y_0) \\
& + \delta^{22}_{1(y+1)/4}(x - x_0)(x - x_1)(y - y_0)(y - y_1) \\
& + \delta^{13}_{(1/2)(y+5)/8}(x - x_0)(y - y_0)(y - y_1)(y - y_2) \\
& + \delta^{04}_{0(y+17)/16}(y - y_0)(y - y_1)(y - y_2)(y - y_3).
\end{aligned} \tag{9}
$$

We limit ourselves to these terms, since only the first terms are ordinarily important, and since the construction of the formula from the terms already written out follows without difficulty. The remainder term can be transformed by use of the mean values of the partial derivatives of corresponding order, as was done in 8(25). Furthermore, it is now evident how the formula will be constructed for more than two variables. We also omit this generalization.

4. *Example*: Atzler[1] gives the energy consumption of a person in lifting various weights from an initial level (e.g., 0 m.) to different heights. The energy consumption in gram calories per kilogram meter of work done is given in the following table:

Weight lifted in kg. / Height in cm.	9.15	13.85	18.95	24.05	28.56
50	76.78	57.55	47.92	42.64	42.87
100	59.30	47.25	40.99	39.26	42.10
150	48.31	39.02	36.07	36.87	39.68
200	44.47	37.26	34.93	36.56	40.69

If the difference scheme is constructed for these values, starting from one corner, and the interpolation formula is set up, we obtain:

$$\begin{aligned} y = 76.78 &- 0.3496(x - 50) - 4.092(y - 9.15) \\ &+ 0.001298(x - 50)(x - 100) \\ &+ 0.03055(x - 50)(y - 9.15) \\ &+ 0.2249(y - 9.15)(y - 13.85) \\ &+ 0.000000880(x - 50)(x - 100)(x - 150) \\ &- 0.0001881(x - 50)(x - 100)(y - 9.15) \\ &- 0.001770(x - 50)(y - 9.15)(y - 13.85) \\ &- 0.00948(y - 9.15)(y - 13.85)(y - 18.95) + \cdots . \end{aligned}$$

If the individual function values are calculated by means of this formula, the following errors are obtained These errors become too large for the extreme values, so that the formula can hardly be used beyond the values used for its construction:

cm	9.15 kg	13.85 kg	18.95 kg	24.05 kg	28.56 kg
50	0	0	0	0	− 7.1
100	0	0	0	− 4.5	−19.4
150	0	0	− 4.9	−18.1	−42.0
200	0	−3.7	−15.3	−40.3	−79.1

If a really useful representation is desired, then it is advisable to start out from a value lying in the middle of the table. It is only necessary to rearrange the table so that this latter value is in one corner; for example, we could select the following arrangement:

cm	18.95 kg	24.05 kg	13.85 kg	9.15 kg	28.56 kg
100	40.99	39.26	47.25	59.30	42.10
150	36.07	36.87	39.02	48.31	39.68
50	47.92	42.64	57.55	76.78	42.87
200	34.93	36.56	37.26	44.47	40.69

If the difference scheme is constructed for this arrangement (which corresponds somewhat to the series of the Gauss formula), the following interpolation formula is obtained:

$$\begin{aligned} y = 40.99 &- 0.0984(x - 100) - 0.3392(y - 18.95) \\ &+ 0.000402(x - 100)(x - 150) \\ &+ 0.009922(x - 100)(y - 18.95) \\ &+ 0.0871(y - 18.95)(y - 24.05) \end{aligned}$$

$$+ 0.00000236(x - 100)(x - 150)(x - 50)$$
$$- 0.00004(x - 100)(x - 150)(y - 18.95)$$
$$- 0.000299(x - 100)(y - 18.95)(y - 24.05)$$
$$- 0.003316(y - 18.95)(y - 24.05)(y - 13.85) \cdots .$$

In this case, considerably smaller errors are obtained:

cm	9.15 kg	13.85 kg	18.95 kg	24.05 kg	28.56 kg
50	−1.55	+2.00	0	0	−2.68
100	0	0	0	0	−2.80
150	−0.97	0	0	0	−1.18
200	−3.36	−3.75	0	+0.68	−1.12

These are errors which barely exceed the limits of accuracy of the measurements. The difference between the two formulas is caused by the fact that the factors with which the divided differences are to be multiplied increase very rapidly in the first arrangement. This behavior is similar to that of the corresponding interpolation formulas for one variable.

5. The calculations are simplified if tables are available with *equidistant function values*. If the difference between the individual values of x is h, and that between the values of y is k, then, if the function values are arranged in order of increasing argument, we have

$$x_1 = x_0 + h; \qquad y_1 = y_0 + k,$$

$$x_2 = x_0 + 2h; \qquad y_2 = y_0 + 2k.$$

If the ordinary differences Δ are introduced in place of the divided differences δ, where

$$\Delta^{10}_{(1/2)0} = z_{10} - z_{00}\,, \qquad \Delta^{01}_{0(1/2)} = z_{01} - z_{00} \qquad \text{etc.,} \tag{10}$$

the general interpolation formula (8) goes over into a formula which corresponds to the special Newton interpolation formula N_+ :[2]

$$N(u, v) = z_{00} + \Delta^{10}_{(1/2)0}u + \Delta^{01}_{0(1/2)}\cdot v + \frac{1}{2!}(\Delta^{20}_{10}u(u - 1)$$
$$+ 2\Delta^{11}_{(1/2)(1/2)}u\cdot v$$
$$+ \Delta^{02}_{01}v(v - 1)) + \frac{1}{3!}(\Delta^{30}_{(3/2)0}u(u - 1)(u - 2) \tag{11}$$

$$+ 3\Delta^{21}_{1(1/2)}u(u-1)v$$

$$+ 3\Delta^{12}_{(1/2)1}uv(v-1) + \Delta^{03}_{0(3/2)}v(v-1)(v-2)) \cdots + R_n^{1})$$

in which u and v have been introduced as new variables,

$$u = \frac{x - x_0}{h}, \qquad v = \frac{y - y_0}{k}.$$

Corresponding to the formulas N_+ and N_- in 10(11) and 10(13), four different formulas can be formed here according to the position of the differences used. For the reasons given above, these formulas are not suitable for the representation of observations, and for interpolation, because differences are used which lie entirely on one side of the initial value of the function. It is always advisable to use differences which are formed from values some of which lie on either side of the function value. This is the case with the Gauss formulas for functions of one variable.

While in the previous treatment (Art. 10) there were only two possible Gauss formulas with the same function value, there are four here, just as with Newton's formula. More generally, with functions of n variables, there are $2n$ different *Gauss formulas.* The first terms of the first formula for functions of two variables are

$$\begin{aligned} z = z_{00} &+ \Delta^{10}_{(1/2)0}u + \Delta^{01}_{0(1/2)}v + \frac{1}{2!}(\Delta^{20}_{00}u(u-1) + 2\Delta^{11}_{(1/2)(1/2)}u \cdot v \\ &+ \Delta^{02}_{00}v(v-1)) \\ &+ \frac{1}{3!}(\Delta^{30}_{(1/2)0}u(u^2-1) + 3\Delta^{21}_{0(1/2)}u(u-1)v + 3\Delta^{12}_{(1/2)0}uv(v-1) \\ &+ \Delta^{03}_{0(1/2)}v(v^2-1)) \\ &+ \frac{1}{4!}(\Delta^{40}_{00}u(u^2-1)(u-2) + 4\Delta^{31}_{(1/2)(1/2)}u(u^2-1)v \\ &+ 6\Delta^{22}_{00}u(u-1)v(v-1) \\ &+ 4\Delta^{13}_{(1/2)(1/2)}uv(v^2-1) + \Delta^{04}_{00}v(v^2-1)(v-2)) + \cdots + R_n\,. \end{aligned} \tag{12}$$

The other formulas are constructed in a similar way. If the mean value is formed from these four formulas, a formula is obtained which corresponds to *Stirling's formula.* This formula is especially useful for interpolation in the vicinity of a point, because the values of the products are small in comparison to the differences. For the construction of the mean value formula, we observe the following. Sums of the four differences appear, and the following abbreviating symbols are used:

$$
(13) \quad
\begin{aligned}
&\Delta^{11}_{(1/2)(1/2)} + \Delta^{11}_{-(1/2)(1/2)} + \Delta^{11}_{(1/2)-(1/2)} + \Delta^{11}_{-(1/2)-(1/2)} = \overline{\Delta^{11}_{00}}\,, \\
&\Delta^{31}_{(1/2)(1/2)} + \Delta^{31}_{-(1/2)(1/2)} + \Delta^{31}_{(1/2)-(1/2)} + \Delta^{31}_{-(1/2)-(1/2)} = \overline{\Delta^{31}_{00}}
\end{aligned}
$$

etc., where, for example, $\overline{\Delta^{11}_{00}}$ is the corresponding difference for the double mesh interval. If this is kept in mind, we obtain the following formula:

$$
\begin{aligned}
S(u, v) &= z_{00} \\
&+ \frac{\Delta^{10}_{(1/2)0} + \Delta^{01}_{-(1/2)0}}{2} u + \frac{\Delta^{01}_{0(1/2)} + \Delta^{01}_{0-(1/2)}}{2} v + \frac{1}{2!} (\Delta^{20}_{00} u^2 + 2\overline{\Delta^{11}_{00}} uv \\
&+ \Delta^{02}_{00} v^2) + \frac{1}{3!}\left[\frac{\Delta^{30}_{(1/2)0} + \Delta^{30}_{-(1/2)0}}{2} u(u^2 - 1) + 3\,\frac{\Delta^{21}_{0(1/2)} + \Delta^{21}_{0-(1/2)}}{2} u^2 v \right.\\
&\quad \left. + 3\,\frac{\Delta^{12}_{(1/2)0} + \Delta^{12}_{-(1/2)0}}{2} vu^2 + \frac{\Delta^{03}_{0(1/2)} + \Delta^{03}_{0-(1/2)}}{2} v(v^2 - 1)\right] \\
&+ \frac{1}{4!} [\Delta^{40}_{00} u^2(u^2 - 1) + 4\overline{\Delta^{31}_{00}} uv(u^2 - 1) + 6\Delta^{22}_{00} u^2 v^2 \\
&\quad + 4\overline{\Delta^{13}_{00}} uv(v^2 - 1) + \Delta^{04}_{00} v^2(v^2 - 1)] + \cdots \\
&+ \frac{1}{5!}\left[\frac{\Delta^{50}_{(1/2)0} + \Delta^{50}_{-(1/2)0}}{2} u(u^2 - 1)(u^2 - 4) + 5\,\frac{\Delta^{41}_{0(1/2)} + \Delta^{41}_{0-(1/2)}}{2} u^2(u^2 - 1)v \right.\\
&\quad + 10\,\frac{\Delta^{32}_{(1/2)0} + \Delta^{32}_{-(1/2)0}}{2} u(u^2 - 1)v^2 + 10\,\frac{\Delta^{23}_{0(1/2)} + \Delta^{23}_{0-(1/2)}}{2} u^2 v(v^2 - 1) \\
&\quad \left. + 5\,\frac{\Delta^{14}_{(1/2)0} + \Delta^{14}_{-(1/2)0}}{2} uv^2(v^2 - 1) + \frac{\Delta^{05}_{0(1/2)} + \Delta^{05}_{0-(1/2)}}{2} v(v^2 - 1)(v^2 - 4)\right] \\
&+ \frac{1}{6!} [\Delta^{60}_{00} u^2(u^2 - 1)(u^2 - 4) + 6\overline{\Delta^{51}_{00}} u(u^2 - 1)(u^2 - 4)v \\
&\quad + 15\Delta^{42}_{00} u^2 \cdot v^2(u^2 - 1) + 20\overline{\Delta^{33}_{00}} u(u^2 - 1)v(v^2 - 1) \\
&\quad + 15\Delta^{24}_{00} u^2 v^2(v^2 - 1) + 6\overline{\Delta^{15}_{00}} uv(v^2 - 1)(v^2 - 4) \\
&\quad + \Delta^{06}_{00} v^2(v^2 - 1)(v^2 - 4)] + \cdots + R_{2n}\,.
\end{aligned}
\tag{14}
$$

6. Finally, the *Bessel formula* can also be written out. We start from the four formulas which are constructed corresponding to the four Gauss formulas above, only that we begin from four different values of the function. These are the values which lie in the corners of a rectangle. We then use only differences on the sides or in the interior of this rectangle. If the mean of these four values is taken and is written as above,

$$
(15) \qquad \Delta_{00} + \Delta_{01} + \Delta_{10} + \Delta_{11} = \overline{4\Delta_{(1/2)(1/2)}}
$$

etc., we obtain the generalized Bessel formula:

$$
\begin{aligned}
B(u,v) = {} & \overline{z_{(1/2)(1/2)}} \\
& + \frac{\Delta^{10}_{(1/2)1}+\Delta^{10}_{(1/2)0}}{2}\left(u-\frac{1}{2}\right)+\frac{\Delta^{01}_{1(1/2)}+\Delta^{01}_{0(1/2)}}{2}\left(v-\frac{1}{2}\right) \\
& + \frac{1}{2!}\left[\overline{\Delta^{20}_{(1/2)(1/2)}}u(u-1)+2\Delta^{11}_{(1/2)(1/2)}\left(u-\frac{1}{2}\right)\left(v-\frac{1}{2}\right)+\overline{\Delta^{02}_{(1/2)(1/2)}}v(v-1)\right] \\
& + \frac{1}{3!}\left[\frac{\Delta^{30}_{(1/2)1}+\Delta^{30}_{(1/2)0}}{2}u(u-1)\left(u-\frac{1}{2}\right)+3\frac{\Delta^{21}_{1(1/2)}+\Delta^{21}_{0(1/2)}}{2}u(u-1)\left(v-\frac{1}{2}\right)\right. \\
& \qquad \left. +3\frac{\Delta^{12}_{(1/2)1}+\Delta^{12}_{(1/2)0}}{2}v(v-1)\left(u-\frac{1}{2}\right)+\frac{\Delta^{03}_{1(1/2)}+\Delta^{03}_{0(1/2)}}{2}v(v-1)\left(v-\frac{1}{2}\right)\right] \\
& + \frac{1}{4!}\left[\overline{\Delta^{40}_{(1/2)(1/2)}}u(u^2-1)(u-2)+4\Delta^{31}_{(1/2)(1/2)}u(u-1)\left(u-\frac{1}{2}\right)\left(v-\frac{1}{2}\right)\right. \\
& \qquad +6\overline{\Delta^{22}_{(1/2)(1/2)}}u(u-1)v(v-1)+4\Delta^{31}_{(1/2)(1/2)}\left(u-\frac{1}{2}\right)v(v-1)\left(v-\frac{1}{2}\right) \\
& \qquad \left. +\overline{\Delta^{04}_{(1/2)(1/2)}}v(v^2-1)(v-2)\right]\cdots+R.
\end{aligned}
\tag{16}
$$

Still other formulas could be derived, but these three will suffice. These formulas have remainder terms which can be expressed by mean values of the partial derivatives. All the considerations with regard to length of interval, remainder terms, etc., which we had discussed for the interpolation series of functions of one variable, can be repeated here.

7. *Example*: In the *Elecktrotechniker-Kalender* the sag of a 35 mm² copper wire is given in centimeters as a function of the length in meters and the temperature in degrees centigrade. The data are reproduced in the following table:

	40 m	60 m	80 m	100 m	120 m	140 m	160 m	180 m	200 m	220 m	240 m
−20°	12	29	63	116	207	310	434	577	744	910	1059
−10°	14	33	71	128	222	329	453	592	760	926	1095
0°	16	38	82	142	240	345	469	609	775	940	1105
+10°	20	45	93	155	252	360	486	624	790	956	1120
+20°	24	53	106	169	266	374	500	640	800	970	1135
+30°	29	62	118	185	282	388	514	659	815	985	1150
+40°	37	70	130	195	292	403	530	675	827	1000	1165

If we use the underscored values, the following difference scheme may be constructed:

		$\underline{116}$	$\Delta^{01}_{-1(1/2)} = 318$	$\underline{434}$		
		$\Delta^{10}_{-(1/2)0} = 26$	$\Delta^{11}_{-(1/2)(1/2)} = 9$	$\Delta^{10}_{-(1/2)1} = 35$		
$\underline{16}$	$\Delta^{01}_{0-(1/2)}$ $= 126$	$\underline{142}$ $\Delta^{02}_{00} = 201$ $\Delta^{20}_{00} = 1$	$\Delta^{01}_{0(1/2)} = 327$ $\Delta^{03}_{0(1/2)} = -57$ $\Delta^{21}_{0(1/2)} = -5$	$\underline{469}$ $\Delta^{02}_{01} = 144$ $\Delta^{20}_{01} = -4$	$\Delta^{01}_{0(1/2)}$ $= 471$	$\underline{940}$
$\Delta^{10}_{(1/2)-1}$ $= 8$	$\Delta^{11}_{(1/2)-(1/2)}$ $= 19$	$\Delta^{10}_{(1/2)0} = 27$ $\Delta^{12}_{(1/2)0} = -15$ $\Delta^{30}_{(1/2)0} = -2$	$\Delta^{11}_{(1/2)(1/2)} = 4$	$\Delta^{10}_{(1/2)1} = 31$ $\Delta^{12}_{(1/2)1} = -5$ $\Delta^{30}_{(1/2)1} = 3$	$\Delta^{11}_{(1/2)(3/2)}$ $= -1$	$\Delta^{10}_{(3/2)2}$ $= 30$
$\underline{24}$	$\Delta^{01}_{1-(1/2)}$ $= 145$	$\underline{169}$ $\Delta^{02}_{10} = 186$ $\Delta^{20}_{10} = -1$	$\Delta^{01}_{(1/2)1} = 331$ $\Delta^{03}_{0(1/2)} = -47$ $\Delta^{21}_{1(1/2)} = 0$	$\underline{500}$ $\Delta^{02}_{11} = 139$ $\Delta^{20}_{11} = -1$	$\Delta^{01}_{1(3/2)}$ $= 470$	$\underline{970}$
		$\Delta^{10}_{(3/2)0} = 26$	$\Delta^{11}_{(3/2)(1/2)} = 4$	$\Delta^{10}_{(3/2)1} = 30$		
		$\underline{195}$	$\Delta^{01}_{2(1/2)} = 335$	$\underline{530}$		

If the generalized Bessel interpolation formula is applied, the following representation is obtained:

$$\begin{aligned} z = 320 &+ 29\left(u - \frac{1}{2}\right) + 329\left(v - \frac{1}{2}\right) - 0.625u(u - 1) \\ &+ 4\left(u - \frac{1}{2}\right)\left(v - \frac{1}{2}\right) + 83.75v(v - 1) \\ &+ 0.0833u(u - 1)\left(u - \frac{1}{2}\right) - 1.25u(u - 1)\left(v - \frac{1}{2}\right) \\ &- 5\left(u - \frac{1}{2}\right)v(v - 1) - 8.667v(v - 1)\left(v - \frac{1}{2}\right), \end{aligned}$$

and if the individual values of the function are calculated with this formula, the deviations obtained from the numbers in the original table are:

		40 m	60 m	80 m	100 m	120 m	140 m
	v	−1	−2/3	−1/3	0	1/3	2/3
	u						
−20°	−1	−9.25	−15.29	−11.32	−1.25	−6.01	−1.54
−10°	−1/2	−6.12	−10.76	− 7.93	+0.44	−5.58	−3.92
0°	0	−2.50	− 6.93	− 7.46	0	−8.49	−3.85
10°	1/2	−0.31	− 4.75	− 6.84	+0.50	−5.66	−3.25
20°	1	+2.50	− 3.15	− 8.01	0	−5.04	−2.07
30°	3/2	+5.00	− 2.08	− 7.91	−2.44	−6.57	−1.24
40°	2	+5.50	+ 0.54	− 7.48	+1.25	−2.18	−1.71

		160 m	180 m	200 m	220 m	240 m	
	v	1	4/3	5/3	2	7/3	
	u						
−20°	−1	+1.25	+2.43	−4.93	+ 2.25	+38.04	
−10°	−1/2	−0.50	+4.76	−4.08	+ 2.05	+16.26	
0°	0	0	+4.12	−3.40	+ 2.50	+18.90	
10°	1/2	−1.19	+4.60	−3.92	− 0.37	+15.02	
20°	1	0	+3.24	−0.92	− 2.50	+ 9.68	
30°	3/2	+0.63	−1.90	−2.74	− 6.83	+ 2.84	
40°	2	−1.25	−4.74	−3.10	−12.25	− 5.13	

Of course, a representation could have been chosen by means of the Stirling formula, starting with a length of 140 m. and a temperature of 10° C.

8. Approximate values for the partial derivatives and for the integral

can be formed for *differentiation* as well as for *integration*, under the same assumptions as in the case of the interpolation formulas for one variable. This is a straightforward operation. The approximate values of the partial derivatives for the argument values x, y for which the values of the function are given, can be computed from Stirling's formula. The values for the points in the interior of the individual rectangles can be obtained from Bessel's formula.

Here we shall write out the formulas for the integrals which use only the last two interpolation formulas. First we shall obtain an approximate value from the *Bessel formula* for the integral over the rectangular region the sides of which have the magnitude of the argument interval.

$$J = \int_{x_0}^{x_0+h} \int_{y_0}^{y_0+k} f(x, y)\,dx\,dy = h \cdot k \int_0^1 \int_0^1 B(u, v)\,du\,dv.$$

$$\begin{aligned} &= h \cdot k \Bigg[\bar{z}_{(1/2)(1/2)} + \frac{1}{2!}\left(\overline{\Delta^{20}_{(1/2)(1/2)}}\left(-\frac{1}{6}\right) + \overline{\Delta^{02}_{(1/2)(1/2)}}\left(-\frac{1}{6}\right)\right) \\ (17) \qquad &+ \frac{1}{4!}\left(\overline{\Delta^{40}_{(1/2)(1/2)}}\,\frac{11}{30} + \frac{1}{6}\,\overline{\Delta^{22}_{(1/2)(1/2)}} + \overline{\Delta^{04}_{(1/2)(1/2)}}\,\frac{11}{30}\right) \\ &+ \frac{1}{6!}\left(\overline{\Delta^{60}_{(1/2)(1/2)}}\left(-\frac{191}{84}\right) + \overline{\Delta^{42}_{(1/2)(1/2)}}\left(-\frac{11}{12}\right)\right. \\ &\left. + \overline{\Delta^{24}_{(1/2)(1/2)}}\left(-\frac{11}{12}\right) + \overline{\Delta^{06}_{(1/2)(1/2)}}\left(-\frac{191}{84}\right)\right) \cdots \Bigg] + R_{2n}\,. \end{aligned}$$

Like the ordinary Bessel integration formula, this formula uses only mean values, i.e., the mean values which are formed from the four values of the function lying at the corners of the rectangle. The similarity to the ordinary Bessel formula becomes more noticeable if we write

$$\begin{aligned} J = h \cdot k \Bigg[&\bar{z}_{(1/2)(1/2)} - \frac{1}{12}\left(\overline{\Delta^{20}_{(1/2)(1/2)}} + \overline{\Delta^{02}_{(1/2)(1/2)}}\right) \\ &+ \frac{11}{720}\left(\overline{\Delta^{40}_{(1/2)(1/2)}} + \frac{5}{11}\,\overline{\Delta^{22}_{(1/2)(1/2)}} + \overline{\Delta^{04}_{(1/2)(1/2)}}\right) \\ (18) \qquad &- \frac{191}{60480}\left(\overline{\Delta^{60}_{(1/2)(1/2)}} + \frac{77}{191}\left(\overline{\Delta^{42}_{(1/2)(1/2)}} + \overline{\Delta^{24}_{(1/2)(1/2)}}\right)\right. \\ &\left. + \overline{\Delta^{06}_{(1/2)(1/2)}}\right) \cdots \Bigg]. \end{aligned}$$

Such formulas can be constructed for a series of adjacent rectangles by use of the corresponding differences, and the results can be summed.

9. If we start with Stirling's formula and integrate over the double interval as in 12.5, we obtain

$$
\begin{aligned}
\int_{x_0-h}^{x_0+h}\int_{y_0-k}^{y_0+k} f(x, y)\,dx\,dy &= h\cdot k\int_{-1}^{+1}\int_{-1}^{+1} S(u, v)\,du\,dv \\
&= h\cdot k\Big[4z_{00} + \frac{1}{2!}\cdot\frac{4}{3}(\Delta_{00}^{20} + \Delta_{00}^{02}) \\
&\quad + \frac{1}{4!}\left(\Delta_{00}^{40}\left(-\frac{8}{15}\right) + \Delta_{00}^{22}\frac{8}{3} + \Delta_{00}^{04}\left(-\frac{8}{15}\right)\right) \\
&\quad + \frac{1}{6!}\left(\Delta_{00}^{60}\cdot\frac{40}{21} - \frac{8}{3}(\Delta_{00}^{42} + \Delta_{00}^{24}) + \Delta_{00}^{06}\frac{40}{21}\right) + \cdots + R \\
&= 4h\cdot k\Big[z_{00} + \frac{1}{6}(\Delta_{00}^{20} + \Delta_{00}^{02}) - \frac{1}{180}(\Delta_{00}^{40} - 5\Delta_{00}^{22} + \Delta_{00}^{04}) \\
&\quad + \frac{1}{1512}\left(\Delta_{00}^{60} - \frac{7}{5}(\Delta_{00}^{42} + \Delta_{00}^{24}) + \Delta_{00}^{06}\right)\cdots\Big] + R_{2n}\,.
\end{aligned}
\tag{19}
$$

Just as in the formula for one variable, only values of the function and differences appear here which are already in the difference scheme. The remainder terms can be formed and estimated as was done in Art. 12 for one variable. A different region of integration, for example a circle, could also have been chosen, but this will not be considered here.

If the Stirling formula is terminated after the second term, a formula corresponding to that of Kepler [12(12)] is obtained for the cubature:

$$
y = 2h\cdot 2k\cdot\frac{z_{01} + z_{10} + z_{0-1} + z_{-10} + 2z_{00}}{6}. \tag{20}
$$

This is also known as *Woolley's formula.*[3] The error in this formula is of fourth order and can be obtained from the Stirling formula. By summing over the various rectangles, a formula may be obtained which corresponds to Simpson's rule, but in which double sums appear. This form can be used for finding the approximate volume of bodies of arbitrary shape.

The formulas of the last sections may also be extended to functions of more than two variables. In this way we can obtain a remarkably simple formula for the approximation of a volume integral, which corresponds to Kepler's barrel rule:

$$y = 2h \cdot 2k \cdot 2l \frac{z_{-100} + z_{100} + z_{0-10} + z_{010} + z_{00-1} + z_{001}}{6}. \tag{21}$$

There appear in this formula only those values of the function which are located in the middle of the bounding surfaces of the body. The error here is also of the fourth order and can be estimated by the generalized formula (19).

Construction of other formulas, such as formulas for multiple integration, are omitted here.

NOTES

1. Atzler, *Naturwissenschaften* (1924), p. 1043.
2. Lambert, *Beiträge*, Teil III (Berlin, 1772); Lagrange, *Nouv. Mém. de Berlin* (1772).
3. Woolley, *Mechanics' Magazine* (1851), p. 262.

CHAPTER THREE

APPROXIMATE INTEGRATION AND DIFFERENTIATION

14. Graphical Methods.

1. Approximation methods are used for integration in all cases in which the functional relation between two variables is not given in the form of an analytical function, but as correlated numerical values, or by a curve. Such methods are also employed when an integration is performed on new functions, not previously tabulated. We have already mentioned (in the preceding chapter) approximation methods which follow directly from the interpolation series. Here, graphical and mechanical, as well as the so-called mean value methods will be discussed. First, we shall consider graphical methods. We therefore assume that the function is given in the form of a curve drawn, perhaps, on rectangular coordinate paper. If we should integrate this curve, it would mean that for the given curve $y = f(x)$ we would find another curve, $Y = F(x)$, the "Integral curve". The ordinates of the curve are a measure of the magnitude of the area which is included between the given curve and the x axis. We consider a rectangle, one side of which is a given length b, the "integration base", and the other is the difference between two ordinates of the integral curve which correspond to the abscissas a and ξ. Then the area of this rectangle,

$$b(Y(\xi) - Y(a)) = \int_a^\xi f(x)\,dx, \tag{1}$$

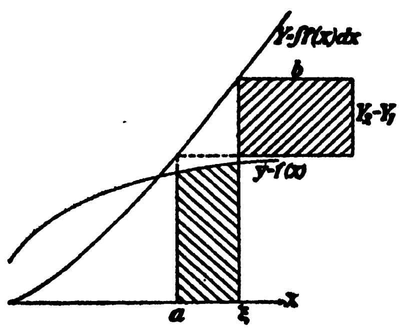


Fig. 34

should be equal to the area of the surface which is bounded by the given curve, the x axis, and the y-parallels through $x = a$ and $x = \xi$. (Both

regions are shaded in Fig. 34.) Since this construction determines only the difference of any ordinates of the integral curve, there are an infinite number of integral curves, all of which may be obtained by the displacement of one such curve in the Y direction. It is therefore sufficient to construct one of these curves, e.g., the one going through the origin. If the integral curve should pass through a given point, the construction is begun at this point and extended in both directions.

2. If, in Fig. 34, we integrate not to the point ξ, but to a neighboring point $\xi + h$, we get the relation

$$b(Y(\xi + h) - Y(a)) = \int_a^{\xi+h} f(x)\,dx. \tag{2}$$

If the relation given in (1) is subtracted from this equation, we have, for a continuous curve,

$$b(Y(\xi + h) - Y(\xi)) = \int_\xi^{\xi+h} f(x)\,dx = h\cdot\bar{y}; \tag{3}$$

$\bar{y}$ designating a *mean ordinate.* Therefore, we have

$$\frac{Y(\xi + h) - Y(\xi)}{h} = \frac{\bar{y}}{b}. \tag{4}$$

If h is allowed to approach zero, then

$$Y'(\xi) = \frac{y}{b}, \tag{5}$$

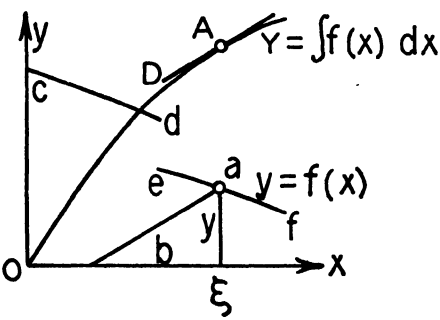


Fig. 35

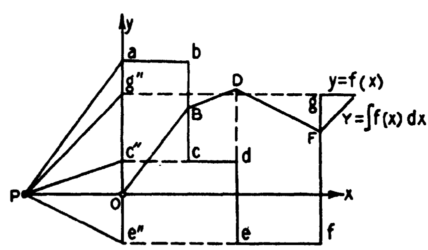


Fig. 36

i.e., *at the points at which the given curve is continuous, it is the derivative curve of the desired integral curve.* The hypotenuse of a triangle, the legs of which are the ordinates of the given curve and the integration base,

has the direction of the tangent of the integral curve at the same value of x. At the point where the given curve has a discontinuity, the integral curve fails to have a tangent (Fig. 35).

3. The integration curve for a straight line parallel to the x axis can easily be constructed. In Fig. 36, the integration base $b = OP$ is drawn from the origin along the negative x axis, and the point a, in which the straight line ab to be integrated cuts the y axis, is connected to the "*integration pole*", P. Then the *direction line Pa* is obtained. This determines the direction of the integral curve. To make the integral curve pass through the origin, we draw a line parallel to Pa passing through the origin.

In the same way, the integral curve $OBDF$ can easily be constructed from the *step function abcdefg* $\cdots$ in Fig. 36. The integral curve is formed from straight lines the slopes of which are determined by the height of the respective steps. It is best to project these ordinates on the y axis by means of lines parallel to the x axis. By connecting these points with the integration pole P, the *pencil of direction lines* (shown in Fig. 36) is obtained. A polygon is then drawn from the origin, corresponding to the individual rays of the pencil. Each side of this polygon is parallel to the direction line of the pencil belonging to the corresponding segment of the original function.

The construction of the pencil of direction lines can be avoided if *auxiliary devices,* such as the integration triangle of Thaer[1] or the integrator of Naatz and Blochmann[2] are used.

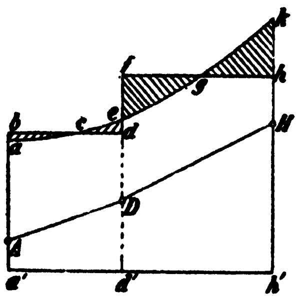


FIG. 37

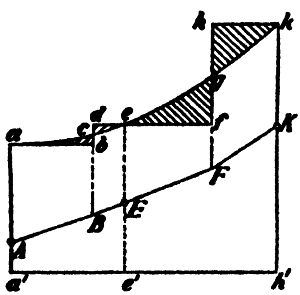


FIG. 38

4. If the integral curve is to be drawn for an arbitrary curve a, c, e, g, k (Fig. 37), the original curve is replaced by a step function with an equal area under the curve. The substitute integral curve, consisting of straight line segments, is drawn for this step function. This gives an approximation for the path of the true integral curve. The substitution of

the given curve can be carried out either by use of *mean ordinates* (Fig. 37) or by *mean abscissas* (Fig. 38). The integration region is divided into strips parallel to the y axis, drawn at appropriate intervals. In curves of slight curvature, the strips may be wide, being narrower for steeper curves. Also, boundary lines of such strips are always drawn through all maxima and minima of the curve. In the first case, the nearly trapezoidal figures $aa'd'e$, etc., with the x axis as one base, and a portion of the curve as the other, are replaced by the rectangles $ba'd'd$, etc. Then the curve segment ace is replaced by an x-parallel bd, so that the triangles abc and cde, bounded by the x-parallel, the curve itself, and the y-parallels bounding the strip, have equal areas. In general, this is done by estimation (cf. Sec. 6).

The step function has then the same area as the given curve up to the boundary of the strip. But the ordinates here are different. Therefore the substitute integral curve and the desired integral curve will have the same ordinates at these limits, but not the same slope. The substitute integral curve is an inscribed polygon of the desired curve. This method is to be recommended for a case in which only the total area is desired.

5. On the other hand, if it is a question of the construction of the integral curve along the entire path, then the method of *mean abscissas* is to be preferred. Parallels to the x axis are drawn through the intersection of the y-parallels bounding the strips and the curve. Another y-parallel is drawn in between each pair of these y-parallels, so that the small triangles abc and cde, lying on either side, are of equal area. The same procedure is repeated for the next interval, giving triangles efg and ghk of equal area, etc. The abscissas of these y-parallels are known as the mean abscissas. Two steps then occur in each strip. The initial step has the same ordinate as the final step of the preceding interval, while the second step has the same ordinate as the initial step of the next interval. In this case the given curve and the step function have equal areas at the boundaries of each strip, and equal ordinates as well. Therefore the substitute integral curve and the desired integral curve have the same ordinate and the same slope at these points. The substitute integral curve is then a succession of tangents of the desired curve, in which the contact points of the individual tangents occur at the boundaries of the strips, and are therefore known. In such a succession of tangents with given points of contact, a curve can be drawn with greater accuracy than is possible with the inscribed polygon discussed previously. Therefore this second method should generally be tried first.

If an area which is bounded by curves above and below is to be determined, then the upper and lower curves are both replaced by step functions. The slope of the substitute integral curve in a given interval

is determined by the difference of ordinates of the upper and lower step curves in this interval (cf. the construction of the first integral curve in Fig. 47).

6. The beginner frequently makes the mistake of choosing the *strip widths* too small. This is not to be recommended, because inaccuracies can arise when a great number of points of connection occur in the series of straight line segments of the substitute curve. With a little practice, the substitution of the given curve with a step curve, even with greater strip widths, can be carried out with great accuracy. If the drawing is made on rectangular coordinate paper, a check can be made on the equal magnitudes of the triangles of a strip by counting the number of squares lying in each triangle.

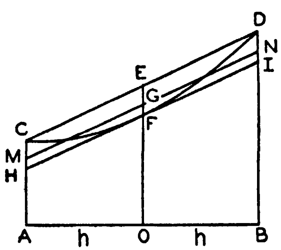


FIG. 39

The following construction gives good results in many cases. Its accuracy corresponds roughly to that of Simpson's rule, and it always is useful if the portion of the curve lying in a given strip can be replaced, with sufficiently good approximation, by a parabola or a cubic.

If the origin of the coordinate system is shifted to the middle of a strip, and the width of the strip is designated by $2h$, then

$$y = a + bx + cx^2 + dx^3, \tag{6}$$

if a cubic is used as a curve. The area of the region $ACFDB$ is then

$$J = \int_{-h}^{+h} y\,dx = 2ah + \frac{2}{3}ch^3. \tag{7}$$

The area of the chord trapezoid $ACDB$ is

$$S = h(y(-h) + y(+h)) = h(2a + 2ch^2), \tag{8}$$

and the area of the tangent trapezoid $AHIB$ is

$$T = 2h \cdot a. \tag{9}$$

From this we see that

$$J = \frac{2T + S}{3}. \tag{10}$$

The distance EF between the middle E of the chord and the contact point F of the tangent parallel to it is divided into three equal parts. If we draw the line MN, parallel to the chord, through the division point G nearest the arc, then the parallelogram $MNDC$ has the same area as

the segment $CDFC$. Consequently, the sides of the parallelogram, CM and DN, can have any direction, provided only that the chord CD and the distance between the two sides CD and MN remain unchanged.

The following construction of the step curve can be given on this basis, using the method of mean abscissas. The strip widths are so chosen that the individual portions of the curve can be approximated by parabolas. Then the chord CD and the tangent parallel to it through F are drawn. Through the middle of the chord E, we draw the x parallel which intersects the tangent at Q. The length EQ is divided into three equal parts. The piece $CKLD$ of the step curve in this strip (which has x-parallels through C and D) is formed by the y-parallel through the division point H adjacent to the tangent.

To justify the correctness of the construction, we observe that in

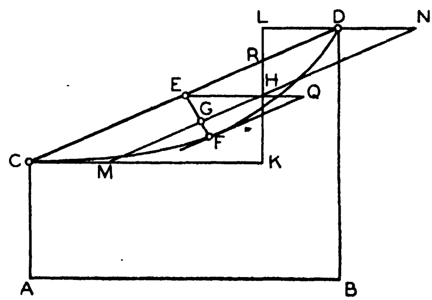


Fig. 40

Fig. 40, instead of subtracting the segment $CFDE$ from the trapezoid $ABDC$, we can subtract the equal area $CMND$. This area can be expressed as

$$CMND = CMHR + \Delta LHN - \Delta LRD$$

$$= CMHR + \Delta MKH - \Delta LRD = \Delta CKR - \Delta LRD.$$

Therefore we can subtract the triangle CKR from the trapezoid, and add the triangle LRD. But this gives the area of the surface bounded by the step curve.

Fig. 41 shows the lines necessary for carrying out the construction. The accuracy of the resultant approximation has been investigated by Thaer.[4] He shows that it is practical to choose the strip width between one fourth and one half of the integration base.

7. Special consideration must be given to the choice of the most desirable *integration base*, and of the *unit* with which the ordinate of the integral curve is to be measured.

The integration base is best chosen so that the integral curve just fills the space at one's disposal, i.e., that the curve has roughly the height of the drawing paper. To achieve this, we first determine approximately

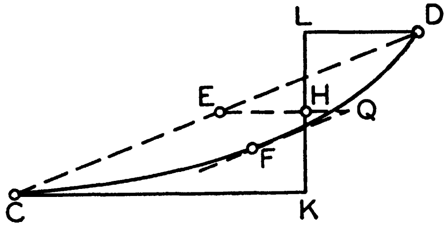


FIG. 41

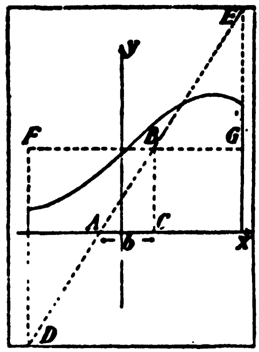


FIG. 42

the mean ordinate of the entire curve. Then we connect the point D, which is the intersection of the y-parallel through the starting point of the interval of integration and the lower margin of the paper, with the point E, which is the intersection of the y-parallel through the end point of the interval of integration and the upper margin of the paper. This line cuts the x axis at A, and the line FG parallel to the x axis through the mean ordinate at B. The distance AC along the x axis between these two points gives a lower limit for the integration base. To get a suitable scale modulus for the ordinate of the integral curve, we choose the smallest integral number of scale divisions greater than or equal to b. This will generally be in centimeters. If the integral curve is then begun at the lower edge of the drawing paper, it will reach to the upper end. The smaller the integration base, the steeper is the curve. If, for any reason, a very steep integral curve is needed, then the integration base is chosen quite small, and the integral curve is drawn in several steps, instead of the usual one. Everytime the upper edge of the paper is reached, a new beginning is made below at the same abscissa, or at some earlier one.

The scale modulus of the ordinate of the integral curve is determined by the following consideration. If the scale modulus of the x axis is E_x cm., and the ordinates of the given curve are plotted with the scale modulus E_y, an increase of the ordinate of the integral curve by one unit corresponds to a rectangle of E_xE_y cm.^{2}, or in E_Y cm., where E_Y is the scale modulus of the integral curve. By (1), this must be

$$bE_Y = E_xE_y .$$

The modulus of the ordinate of the integral curve is then determined by the equation

$$E_Y = \frac{E_x \cdot E_y}{b}. \tag{11}$$

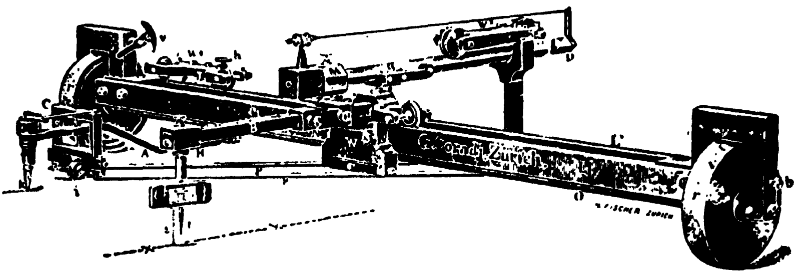


FIG. 43

8. The integral curve can also be drawn mechanically. It is only necessary to have a drawing device which traces a curve, the slope of which

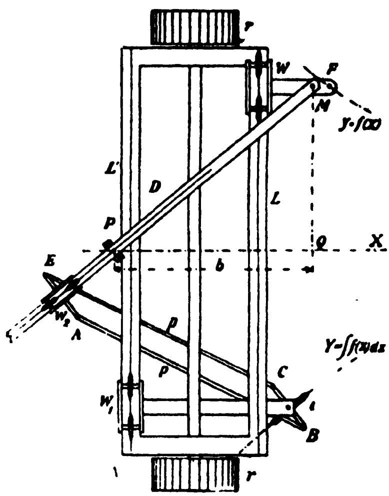


FIG. 44

proportional to the ordinate of the given curve. This device has a ɔinter which is moved by hand over the given curve. Such an instru-ənt is known as an integrator. The most commonly used integraph as constructed by Abdank-Abakanowitz[5] and improved by Napoli and ɔradi. Fig. 44 shows the schematic diagram of this apparatus, while g. 43 shows the Coradi model. It consists of a carriage with two rods, and L', parallel to the y axis, which may be moved in the x direction ʳ means of the rollers r. The slope rod D slides along by means of the

pin P located at the middle of the rod. This rod can be turned about the pivot M. This pivot is found on the differential carriage W, which slides along the rod L by means of the tracing point F. The tracing point F is moved along the given curve by hand. The pivot P moves along the x axis. If the distance PQ, parallel to the x axis, of the two pivots M, P, is denoted by b, then the slope of the slope rod is y/b; the rod is therefore parallel, at each point, to the slope of the desired integral curve. This is accomplished by the sharp-edged roller i, which turns about a pivot of the integral carriage W_1. This carriage slides along the rod L'. A sharp-edged roller can only be moved in its own plane. In the hinged parallelogram $ABCE$, one side is the axis of the roller. The opposite side can so move along the slope rod D (by means of the carriage W_2) that it is always perpendicular to the rod. Therefore the intersection line of the plane of the roller and the paper is always parallel to the slope rod D. Therefore the roller traces a curve which has at each point the prescribed slope. This curve is therefore the integral curve. The integration base b, i.e., the x-parallel distance between the pins P and M, can be adjusted between the values of 10 and 20 cm. By this means, the integral curve is kept, as far as possible, within the range of measurement of the rod.

For adjustment, the differential carriage, which can be clamped at each point of the rod L (which is provided with a millimeter scale) is clamped in the middle of the rod, and the integral carriage, the roller of which can be raised by a screw, is also placed at the middle of L'. The tracing point and the integral roller must then be moved, by means of a shifting of the integrator, along the x axis. The use of the integrator is recommended only if a large number of integrations are to be performed, since a simple graphical integration could easily be carried out in the time required for setting up and adjusting the integrator.

9. *Example*: A radially symmetric vessel, the vertical cross-section of which is drawn in Fig. 45, is to be calibrated. The volume of the vessel up to the height h is

$$y(h) = \int_0^h r^2\pi \, dh, \tag{12}$$

where r is a function of h. To evaluate this integral, we first draw r^2 as a function of h. In this case, the scale of the ordinate is changed, if necessary, so that the new curve fits conveniently on the graph paper. In Fig. 45, which is reduced to one third actual size, one centimeter was chosen as the unit, and the scale was not changed. If this curve is integrated, we obtain another curve, measured with the corresponding unit, which represents the above integral. The

scale modulus of the h axis is $E_x = 1$ cm., while that of the curve to be integrated is $E_y = 1/\pi$, because π was not considered in the drawing. The value $b = 3$ cm. is chosen as the integration base, so that the scale modulus of the ordinate of the integral curve is then $E_y =$

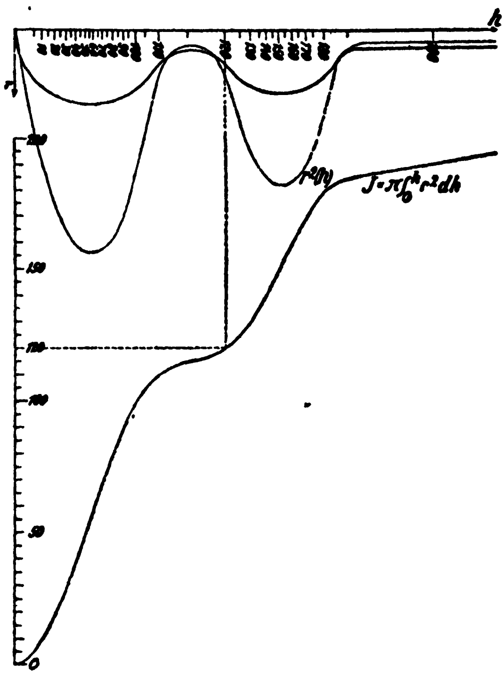


FIG. 45

$1/3\pi$ cm., i.e., 1 cm.^{3} of volume corresponds to $1/3\pi$ cm. of ordinate. Equal ordinate differences correspond to equal volume differences. Therefore, if the end ordinate of the integral curve is divided into equal subintervals of length about 5.3 mm., then the distance between two consecutive scale marks corresponds to a volume change of 5 cc. If we draw a line parallel to the h axis out to the integral curve, as has been done at the mark 120 in the figure, and if this point of the integral curve is projected on the h axis, then the height h in the vessel has been ascertained at which the capacity has the value 120 cm.^{3} A non-linear scale is then obtained, from which we can read the actual volume of the vessel in cubic centimeters.

10. Repeated integration of an integral curve gives the double integral curve; an integration of this yields the triple integral curve, etc. These

multiple integral curves can be used for the evaluation of centers of mass, moments of inertia, and the higher moments of plane pieces and cross sections of prismatic bodies, as well as the centers of mass of solid bodies, particularly of figures of revolution.[6]

The static moment of a plane surface with uniform mass distribution is zero about any line through the center of mass. If it is possible to find two lines, about which the static moment of the plane surface is zero, then

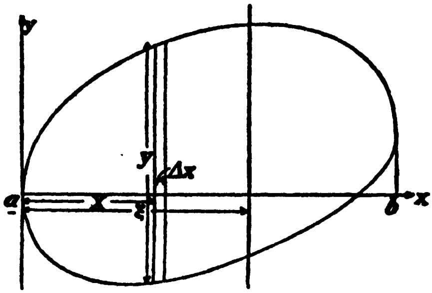


FIG. 46

their intersection locates the center of mass. The moment of a plane surface about a y-parallel at a distance ξ from the y axis is

$$\eta = \int_a^b \int_{y_u}^{y_o} (\xi - x)\,dx\,dy = \int_a^b (\xi - x)y\,dx, \tag{13}$$

where y is the difference of ordinates between the upper and lower boundaries, and where the moment is taken as positive in the counterclockwise direction, looking in the positive direction of the y axis. Integration of this expression by parts gives

$$\eta = \int_a^b (\xi - x)y\,dx = (\xi - b)\int_a^b y\,dx + \int_a^b \int_a^x y\,dx^2. \tag{14}$$

If the definite integrals are designated by $Y_1(b)$ and $Y_2(b)$, we have

$$\eta = Y_2(b) + (\xi - b)Y_1(b). \tag{15}$$

This is the equation of a straight line which has $Y_2(b)$ for the ordinate intercept and $Y_1(b)$ for the slope. The line is therefore tangent to the double integral curve at its endpoint. Therefore, as is shown in Fig. 47, the first integral curve AB is drawn for the closed area *acbd* (the center of mass of which is to be determined) from $x = 0$, $y = 0$ out. Corresponding to this (and shifted upward in Fig. 47) is drawn the double integral curve A_1B_1. The end tangent of this curve, B_1F_1, gives the static moment for the various lines parallel to the x axis. In particular, the moment is zero for a y-parallel passing through the intersection E_1

of F_1B_1 with the x axis. Therefore the center of mass must lie on a line parallel to the y axis through E_1. If the same construction is carried out

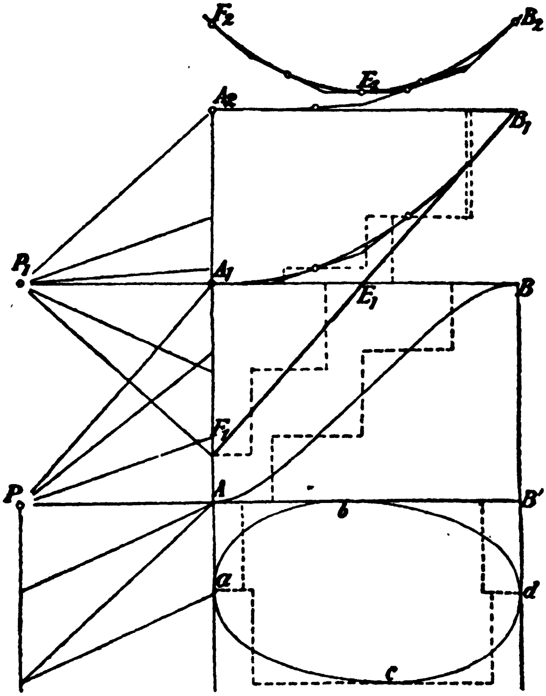


FIG. 47

for the lines parallel to the x axis, a second line is obtained, and the intersection of these lines determines the center of mass.

The moment of inertia about a line parallel to the y axis at a distance ξ is

$$(16) \qquad T = \int_a^b \int_{y_u}^{y_0} (\xi - x)^2 \, dx \, dy = \int_a^b y(\xi - x)^2 \, dx.$$

From this we can obtain

$$(17) \qquad \begin{aligned} T &= (\xi - b)^2 \int_a^b y \, dx + 2 \int_a^b (\xi - x) \int_a^x y \, dx \, dx \\ &= (\xi - b)^2 \int_a^b y \, dx + 2(\xi - b) \int_a^b \int_a^x y \, dx^2 + 2 \int_a^b \int_a^x \int_a^x y \, dx^3. \end{aligned}$$

by a double application of integration by parts. For one half of the moment of inertia, we have

$$\frac{T}{2} = \bar{\eta} = \frac{(\xi - b)^2}{2} Y_1(b) + (\xi - b)Y_2(b) + Y_3(b), \tag{18}$$

with the same abbreviations as before. This is a parabola, which is obtained by integration of the lines determining the static moment, if this parabola is laid through the end point of the triple integral curve. We integrate the second integral curve A_1B_1 once more from the point $x = 0$, $y = 0$ out. We then draw the integral curve $B_2E_2F_2$ of the line $B_1E_1F_1$ from the endpoint B_2 of this curve. The ordinate of this curve, measured in corresponding units, is equal to one half the moment of intertia about the y-parallel with the same abscissa. The moment of inertia about the straight line through the center of gravity is naturally a minimum.

By continuing in this fashion, we can evaluate the moments of higher order. The moment of nth order (n a positive number) leads, as Jacobi[7] has shown, to a curve of nth degree which can be found by repeated use of integration by parts:

$$\begin{aligned} \int_a^b (\xi - x)^n y\,dx &= (\xi - b)^n \int_a^b y\,dx \\ &+ n(\xi - b)^{n-1} \int_a^b \int_a^x y\,dx^2 + \cdots \\ &+ n(n-1)(\xi - b)^{n-2} \int_a^b \int_a^x \int_a^x y\,dx^3 + \cdots \\ &+ n! \int_a^b \int_a^x \cdots \int_a^x y\,dx^{n+1}. \end{aligned} \tag{19}$$

These integral curves find important application in the theory of beams. The first integral curve of the stress curve permits the determination of the transverse force, the second the bending moment, the third the angle of inclination, and the fourth the elastic line, if small sagging can be assumed.[8]

11. The construction of the *differential curve* of a given curve is of at least equal importance to the construction of the integral curve. This method requires the construction of tangents to the given curve. The trigonometric tangent or slope of the curve is then the ordinate of the differential curve at that abscissa. However, the construction of the tangent is possible with only limited accuracy. This is not due to the method, but to the nature of the task. For each line has a certain width, which masks the finer fluctuations which are of importance in the con-

struction of the differential curve. On the other hand, small errors in the drawing of the given curve, such as are due to the inertia or vibrations of the drawing apparatus, to the rubbing of the drawing pencil on the paper, etc., can lead to large fluctuations in the slope, and therefore to completely inaccurate differential curves. We neglect here such inaccuracies as are caused by the method of drawing, and set ourselves the task of drawing the differential curve of an existing curve.

This problem reduces to that of the most accurate construction of the tangents.

a. The following method is given by Lambert[9] for the construction of a tangent of given slope and contact point. He draws a number of chords of prescribed slope in the curve, bisects them, and lays a smooth curve through these midpoints. This curve intersects the given curve at the

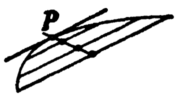


Fig. 48

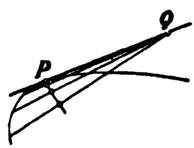


Fig. 49

point of contact P, through which the tangent can then be drawn (Fig. 48).

b. Fig. 49 shows how this construction can be altered if the tangent PQ is drawn to a given portion of the curve from an exterior point Q.[10]

c. Gugler[11] gives the following method for the construction of the tangent PQ to a point P of a curve (Fig. 50). Through the point P is drawn a series of secants which cut the curve for the second time at the

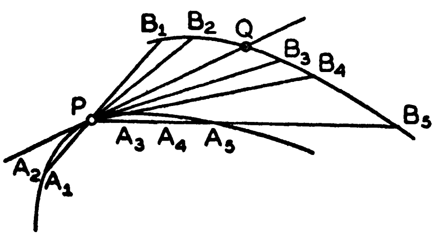


Fig. 50

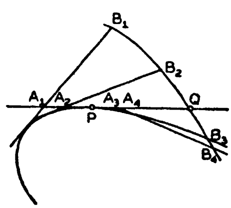


Fig. 51

points A_m. These lines are extended an appropriately chosen distance l beyond these points. The points B_m are then obtained, and are connected by a smooth curve. If we now describe a circle about P of radius l, the curve through the points B_m will be cut by this curve at a point Q, which must then lie on the tangent through P.

d. Mehmke[12] has developed a method from this to find the point of contact of a previously drawn tangent. He draws (approximately) a series of neighboring tangents (Fig. 51). He then marks off equal lengths l from

the points of intersection A_m of these lines with the given tangent. Through the points B_m thus determined, he draws a smooth curve, which cuts the given tangent in Q. If the distance l is laid out from Q toward the curve, the point of intersection P is determined.

Tangents can be constructed more rapidly and also more accurately with several simple devices. The best known of these are the mirror ruler of Reusch, and the tangent drawer of Pfluger.[13]

12. *The drawing of the differential curve* can be carried out if a series of tangents are drawn to the given curve. This is shown by the following: We can draw directly the differential curve which corresponds to the

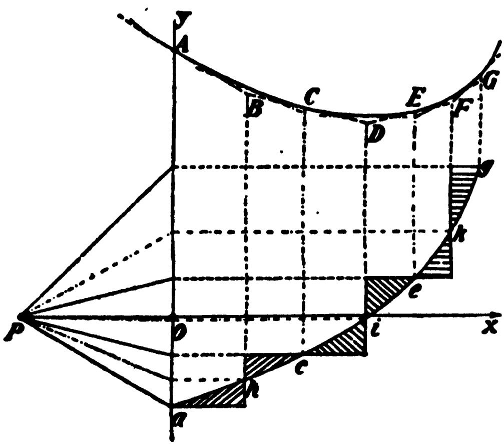


FIG. 52

tangent polygon $ABCDEFG$ (Fig. 52), constructed by the above method. This has the form of a step curve. The height of the steps is found by means of a pencil of lines whose rays are parallel to the corresponding tangents. Also, the points *ace* at which the desired differential curve coincides with the step curve are easily found by vertical lines drawn from the contact points ACE of the tangents. But the drawing of the differential curve itself is difficult. If we reverse the construction of the integral curve by the method of mean abscissas, we find that the desired curve must be drawn so that the triangles, lying adjacent to the same vertical portion of the step curve, are equal. The intersection of the differential curve with this portion of the step curve can be found by constructing the tangent of the given curve at the point with the same abscissa (dotted lines in Fig. 52). The drawing of the differential curve is materially simplified by a determination of this point of intersection.

13. Finally, we shall discuss the *determination of the arc length* of a curve. This is, of course, measured by $\int ds$ and is then approximated by $\sum \Delta s$. To determine the length of the arc of a given curve, we adjust the separation of the points of a compass to a small length d. This length is marked off as a chord, perhaps n times, along the curve. Then, approximately, $s \approx nd$. The difference between the curved arc and the somewhat smaller chords can be compensated if the compass points are set somewhat outside of the curve on the convex side. In addition, special compasses have been constructed which close up each time after measuring off a chord. They then give the sum of the lengths of the chords measured. These have the advantage that the length of chord varied to correspond to greater or lesser curvature.

A measuring wheel can also be used for the measurement of arc length. This is employed in the measurement of distances on maps. Such measuring wheels are not to be recommended for small arc lengths, since the inserting and reading accuracy are rather small. The distance between two markings corresponds in general to the length of 1 cm. For more exact measurement of curve lengths, special devices known as curve meters, are constructed by Amsler, Coradi and others.

To determine the length of a space curve, given in piecewise projection, we consider a cylinder through the curve, perpendicular to the base plane, which is unrolled. The cross section curve is divided up equally by means of the compass, as described above, and this graduation is transferred to a straight line, perhaps the x axis. Then the position of the individual segments is estimated and is projected on the x axis. The heights are plotted perpendicular to the x axis at the given points, in the same scale. The points so located are connected by a smooth curve, and the length of this curve is measured.[14]

NOTES

1. Thaer, *Zeits. f. angew. Math. u. Mech.* 6 (1926), p. 252.
2. Naatz-Blochmann, *Das zeichnerische Integrieren mit dem Integranten* (1921).
3. Cf. 12(12), 15.4.
4. Thaer, *Zeits. f. angew. Math. u. Mech.* 7 (1927), pp. 152-155.
5. Abdank-Abakanowitz, *Die Integraphen* (Leipzig, 1889).
6. Nehls, *Graphische Integration* (Leipzig, 1885), p. 57 ff.
7. Jacobi, *Journal für Mathematik I* (1825), p. 301.
8. Massau, *Intégration graphique* (1878-85). Nehls, *Graphische Integration* (1885). Other applications are given by Gerstenbrandt, *Die graphische Integration* (Wittenberg, 1926).
9. Lambert, *Beiträge II* (Berlin, 1770), 2nd ed., Art. 21.
10. Gergonne, *Annales de Gergonne XII* (1821-22), p. 137.
11. Gugler, *Lehrbuch der deskriptiven Geometrie*, 2nd ed. (Stuttgart, 1857), p. 151.
12. Mehmke, *Leitfaden zum graphischen Rechnen* (Leipzig, 1917), p. 111.
13. Willers, *Math. Instrumente* (Berlin, 1926), Art. 7.

14. Further applications of graphical integration are found in Gerstenbrandt, *op. cit;* Massau, *op. cit;* Rothe, *Elektrotechnische Zeitschrift*, 41 (1920), pp. 999-1002. The application of graphical integration to the determination of the road time of railroad trains is also important. Cf. Nordmann, *Glasers Annalen*, 101 (1927), pp. 163-175.

15. Euler's Formula.

1. For the evaluation of integrals, as also for the rapid calculation of sums of equidistant function values, a formula can be used which was advanced by Euler[1] and probably independently by MacLaurin.[2] This formula gives a value of an integral by a series of function values of the function being integrated, and of derivatives within the limits of the interval. To derive the Euler formula, we start from the integral formula 12(9) which follows from the interpolation formula of Bessel:

$$\begin{aligned}\int_{x_0}^{x_0+\kappa h} f(x)\,dx = h\Big[\sum_0^{\kappa} y_\lambda - \frac{y_0+y_\kappa}{2} - \frac{1}{12}(\overline{\Delta_\kappa^1} - \overline{\Delta_0^1}) \\ + \frac{11}{720}(\overline{\Delta_\kappa^3} - \overline{\Delta_0^3}) - \frac{191}{60480}(\overline{\Delta_\kappa^5} - \overline{\Delta_0^5}) + \frac{2497}{3628800}(\overline{\Delta_\kappa^7} - \overline{\Delta_0^7}) \cdots\end{aligned} \tag{1}$$

with the remainder term

$$R_{2n} = \kappa h^{2n+1} C \cdot f^{(2n)}(\xi), \qquad x_0 - (n-1)h \leqq \xi \leqq x_0 + (n+\kappa)h, \tag{2}$$

where C is the integral over the function of the nth term of the Bessel interpolation series.

In this formula there appear only values of the rows 0 and κ, i.e., mean values of the values appearing in the difference scheme which lie a half row higher and lower. But the same mean values would have been obtained if the derivatives of the function had been formed from the representation by means of the Stirling formula. From 11(6) we find, by continued differentiation,

$$\begin{aligned}h^1 f'(x_0) &= \overline{\Delta_0^1} - \frac{1}{6}\overline{\Delta_0^3} + \frac{1}{30}\overline{\Delta_0^5} - \frac{1}{140}\overline{\Delta_0^7} \cdots, \\ h^3 f'''(x_0) &= \overline{\Delta_0^3} - \frac{1}{4}\overline{\Delta_0^5} + \frac{7}{120}\overline{\Delta_0^7} \cdots, \\ h^5 f^5(x_0) &= \overline{\Delta_0^5} - \frac{1}{3}\overline{\Delta_0^7} \cdots, \\ h^7 f^7(x_0) &= \overline{\Delta_0^7} \cdots,\end{aligned} \tag{3}$$

if we set $t = 0$. Here the function $f(x)$ is inserted, instead of the approxi-

mating function $S(x)$, under the assumption that this is differentiable a corresponding number of times. By step-by-step elimination, we obtain

$$
(4)\qquad
\begin{aligned}
\overline{\Delta_0^7} &= h^7 f^{(7)}(x_0) \cdots ,\\
\overline{\Delta_0^5} &= h^5 f^{(5)}(x_0) + \frac{1}{3} h^7 f^{(7)}(x_0) \cdots ,\\
\overline{\Delta_0^3} &= h^3 f'''(x_0) + \frac{1}{4} h^5 f^{(5)}(x_0) + \frac{1}{40} h^7 f^{(7)}(x_0) \cdots ,\\
\overline{\Delta_0^1} &= h f'(x_0) + \frac{1}{6} h^3 f'''(x_0) + \frac{1}{120} h^5 f^{(5)}(x_0) + \frac{1}{5040} h^7 f^{(7)}(x_0) \cdots .
\end{aligned}
$$

Corresponding values are obtained for the $\overline{\Delta_\kappa}$. If these values are substituted in the given formula, we obtain *Euler's formula*:

$$
(5)\qquad
\begin{aligned}
\int_{x_0}^{x_\kappa} f(x)\,dx = h \sum_0^\kappa y_\lambda - h\frac{y_0 + y_\kappa}{2} - \frac{h^2}{12}(f'(x_\kappa) - f'(x_0))\\
+ \frac{h^4}{720}(f'''(x_\kappa) - f'''(x_0)) - \frac{h^6}{30240}(f^{(5)}(x_\kappa) - f^{(5)}(x_0))\\
+ \frac{h^8}{1209600}(f^{(7)}(x_\kappa) - f^{(7)}(x_0)) \cdots .
\end{aligned}
$$

The coefficient of the nth term is usually written $B_n/n!$ where the B_n are the *Bernoulli numbers*[3]. The factor C of the remainder term will naturally have another value here than in the above formula, but this value will not be derived. We now write out the Euler integral formula for the double interval:

$$
(6)\qquad
\begin{aligned}
\int_{x_0}^{x_{2\kappa}} f(x)\,dx = 2h \sum_0^\kappa y_{2\lambda} - (y_0 + y_{2\kappa})h - \frac{(2h)^2}{12}(f'(x_{2\kappa}) - f('x_0))\\
+ \frac{(2h)^4}{720}(f'''(x_{2\kappa}) - f'''(x_0)) - \frac{(2h)^6}{30240}(f^{(5)}(x_{2\kappa}) - f^{(5)}(x_0)) \cdots .
\end{aligned}
$$

2. From this formula, the form of the expression can be derived which was used by MacLaurin, and which employs only the odd ordinates. We first consider the formula for the interval from 0 to $2h$ with the limit ordinates y_0 and y_2 and mean ordinate y_1. If this function is sufficiently differentiable in the interval in question, it can be expanded about the middle of the interval by Taylor's series, so that

$$
(7)\qquad y_0 + y_2 = 2y_1 + h^2 f''(x_1) + 2\frac{h^4}{4!} f''''(x_1) + \cdots .
$$

If $f'(x)$ is developed in the same way about x_1, and $-h$ and $+h$ are substituted and the subtraction $f'(x_2) - f'(x_0)$ performed, we have

$$\text{(7a)} \quad f'(x_2) - f'(x_0) = 2\left(hf''(x_1) + \frac{h^3}{3!}f''''(x_1) + \frac{h^5}{5!}f^{(6)}(x_1) \cdots\right).$$

If the value obtained in this expression for $f''(x_1)$ is substituted in the expansion for $y_0 + y_2$, we have

$$\text{(8)} \quad \begin{aligned} y_0 + y_2 &= 2y_1 + \frac{h}{2}(f'(x_2) - f'(x_0)) - h^4f^{(4)}(x_1)\left(\frac{1}{3!} - \frac{2}{4!}\right) \\ &\qquad - h^6f^{(6)}(x_1)\left(\frac{1}{5!} - 2\frac{1}{6!}\right) \cdots \\ &= 2y_1 + \frac{h}{2}(f'(x_2) - f'(x_0)) - \frac{2h^4}{4!}f^{(4)}(x_1) - \frac{4h^6}{6!}f^{(6)}(x_1) \\ &\qquad - \frac{6h^8}{8!}f^{(8)}(x_1) \cdots . \end{aligned}$$

If the value of $f^{(4)}(x_1)$ is calculated, just as above,

$$\text{(8a)} \quad f'''(x_2) - f'''(x_0) = 2\left(hf^{(4)}(x_1) + \frac{h^3}{3!}f^{(6)}(x_1) + \frac{h^5}{5!}f^{(8)}(x_1) \cdots\right),$$

and is substituted in (8), we obtain

$$\text{(9)} \quad \begin{aligned} (y_0 + y_2) &= 2y_1 + \frac{h}{2!}(f'(x_2) - f'(x_0)) - \frac{h^3}{4!}(f'''(x_2) - f'''(x_0)) \\ &\qquad + h^6f^{(6)}(x_1)\left(\frac{2}{3!4!} - \frac{4}{6!}\right) + h^8f^{(8)}(x_1)\left(\frac{2}{5!4!} - \frac{6}{8!}\right) \cdots \\ &= 2y_1 + \frac{h}{2!}(f'(x_2) - f'(x_0)) - \frac{h^3}{4!}(f'''(x_2) - f'''(x_0)) \\ &\qquad + \frac{6h^6}{6!}f^{(6)}(x_1) + \frac{22}{8!}h^8f^{(8)}(x_1) \cdots . \end{aligned}$$

If we continue in this way, we get

$$\text{(10)} \quad y_0 + y_2 = 2y_1 + \frac{h}{2}\Delta f' - \frac{h^3}{4!}\Delta f''' + \frac{3h^5}{6!}\Delta f^{(5)} - \frac{17h^7}{8!}\Delta f^{(7)} \cdots ,$$

where $\Delta f^\kappa = f^{(\kappa)}(x_2) - f^{(\kappa)}(x_0)$. Now, if we take the formula (6), which was written for a double interval, and substitute the value calculated here for $y_0 + y_2$, we get

$$
\begin{aligned}
\int_{x_0}^{x_2} f(x)\,dx &= 2hy_1 + \frac{h^2}{2}\,\Delta f' - \frac{h^4}{24}\,\Delta f''' + \frac{h^6}{240}\,\Delta f^{(5)} - \frac{17h^8}{40320}\,\Delta f^{(7)} \cdots \\
&\quad - \frac{(2h)^2}{12}\,\Delta f' + \frac{(2h)^4}{720}\,\Delta f''' - \frac{(2h)^6}{30240}\,\Delta f^{(5)} + \frac{(2h)^8}{1209600}\,\Delta f^{(7)} \cdots \qquad (11) \\
&= 2hy_1 + \frac{(2h)^2}{24}\,\Delta f' - \frac{7(2h)^4}{5760}\,\Delta f''' + \frac{31(2h)^6}{967680}\,\Delta f^{(5)} - \frac{127(2h)^8}{154828800}\,\Delta f^{(7)} \cdots
\end{aligned}
$$

If we write this formula for the intervals x_2 to x_4 , x_4 to x_6 , etc., and sum, we obtain the *MacLaurin form* of the Euler formula:

$$
\begin{aligned}
\int_{x_0}^{x_{2\kappa}} f(x)\,dx &= 2h \sum_1^{\kappa} y_{2\lambda-1} + \frac{(2h)^2}{24}\,(f'(x_{2\kappa}) - f'(x_0)) \\
&\quad - \frac{7(2h)^4}{5760}\,(f'''(x_{2\kappa}) - f'''(x_0)) + \frac{31(2h)^6}{967680}\,(f^{(5)}(x_{2\kappa}) - f^{(5)}(x_0)) \qquad (12) \\
&\quad - \frac{127(2h)^8}{154828800}\,(f^{(7)}(x_{2\kappa}) - f^{(7)}(x_0)) \cdots .
\end{aligned}
$$

3. By means of the Euler formula, other formulas can be derived for the *sum of the nth power of the integers*, where n is a positive integer.[4] If $f(x) = x^n$ is substituted in the Euler formula (5), there results

$$
\begin{aligned}
\int_0^{\kappa} x^n\,dx = \frac{\kappa^{n+1}}{n+1} &= \sum_0^{\kappa} \lambda^n - \frac{\kappa^n}{2} - \frac{n}{12}\,\kappa^{n-1} \\
&\quad + \frac{n(n-1)(n-2)}{720}\,\kappa^{n-3} \cdots , \qquad (12a)
\end{aligned}
$$

so that

$$
\sum_0^{\kappa} \lambda^n = \frac{\kappa^{n+1}}{n+1} + \frac{\kappa^n}{2} + \frac{n}{12}\,\kappa^{n-1} - \frac{n(n-1)(n-2)}{720}\,\kappa^{n-3} \cdots . \qquad (13)
$$

The last term contains either κ or κ^2 as a factor. From the above formula we obtain, for various n

$$
\begin{aligned}
&n = 1: \frac{\kappa^2 + \kappa}{2}, &\qquad &n = 3: \frac{\kappa^4}{4} + \frac{\kappa^3}{2} + \frac{\kappa^2}{4}, \\
&n = 2: \frac{\kappa^3}{3} + \frac{\kappa^2}{2} + \frac{\kappa}{6}, &\qquad &n = 4: \frac{\kappa^5}{5} + \frac{\kappa^4}{2} + \frac{\kappa^3}{3} - \frac{\kappa}{30}. \qquad (14)
\end{aligned}
$$

Similarly, formulas for the *sum of the nth power of the odd integers* can be developed by means of the MacLaurin formula (12). If we set $f(x) = x^n$, then

$$\begin{aligned}\frac{1}{2}\int_0^{2\kappa} x^n\,dx = \frac{(2\kappa)^{n+1}}{2(n+1)} &= \sum_1^{\kappa} (2\lambda - 1)^n \\ &\quad + \frac{n}{12}(2\kappa)^{n-1} - \frac{7n(n-1)(n-2)}{720}(2\kappa)^{n-3} \cdots,\end{aligned} \tag{14a}$$

or

$$\sum_1^{\kappa} (2\lambda - 1)^n = \frac{(2\kappa)^{n+1}}{2(n+1)} - \frac{n}{12}(2\kappa)^{n-1} + \frac{7n(n-1)(n-2)}{720}(2\kappa)^{n-3}. \tag{15}$$

As special cases, we have the following:

$$\begin{aligned} &n = 1: \kappa^2, && n = 3: 2\kappa^4 - \kappa^2, \\ &n = 2: \frac{4\kappa^3}{3} - \frac{\kappa}{3}, && n = 4: \frac{16}{5}\kappa^5 - \frac{8}{3}\kappa^3 + \frac{7}{15}\kappa. \end{aligned} \tag{16}$$

If the function appearing in the integral of the Euler formula is easy to integrate, we can also use the formula to obtain the sum of equidistant function values. In this case we can also use a formula developed by Lubbock.[5] This does not use all the given summands, but only a small number of equidistant values.

4. In general, the derivatives are difficult to form. Therefore, for approximate calculation of definite integrals, we sometimes use only the terms of this formula containing the values of the function. By combination of the Euler formula for various interval widths and the MacLaurin formula, a series of very well known approximation formulas are obtained, as well as the corresponding correction terms. (cf. Art. 12).

If only the function values of the Euler formula are taken, the so-called trapezoid rule is obtained:

$$J_T = \left(\frac{y_0}{2} + y_1 + y_2 + \cdots + \frac{y_\kappa}{2}\right) h = h\left(\sum_0^{\kappa} y_\lambda - \frac{y_0 + y_\kappa}{2}\right), \tag{17}$$

with the correction term

$$K_T = -\frac{h^2}{12}(f'(x_\kappa) - f'(x_0)) + \frac{h^4}{720}(f'''(x_\kappa) - f'''(x_0)) \cdots. \tag{18}$$

In exactly the same way, the function values of the MacLaurin formula can be used. Then we obtain the *tangent trapezoid rule*:

$$(19)\qquad J_L = (y_1 + y_3 + \cdots + y_{2\nu-1})2h = 2h \sum_0^{\nu} y_{2\lambda-1} ,$$

with the correction

$$(20)\quad K_L = \frac{h^2}{6}(f'(x_{2\nu}) - f'(x_0)) - \frac{7h^4}{360}(f'''(x_{2\nu}) - f'''(x_0)) \cdots .$$

The first term of the correction has the opposite sign in the two formulas. If we set $\kappa = 2\nu$, the term is half as large in the first case as in the second. If we judge the value of the approximation which can be obtained with such a formula by the number of terms of the Euler formula which can be represented by it, then it is obvious that the first term of the correction can be eliminated by a combination of the two formulas. We then obtain *Simpson's rule,*[6] already discussed in 12.5:

$$(21)\qquad J_S = \frac{2J_T + J_L}{3} = \frac{h}{3}\sum_0^{\nu}(y_{2\lambda} + 4y_{2\lambda+1} + y_{2\lambda+2}),$$

with the correction

$$(22)\ K_S = -\frac{h^4}{180}(f'''(x_{2\nu}) - f'''(x_0)) + \frac{h^6}{1512}(f^{(5)}(x_{2\nu}) - f^{(5)}(x_0)).$$

The formula is to be used only if there is an even number of subintervals available; i.e., an odd number of function values. While the width of the subinterval appears as a squared term in both the first two formulas, here it occurs as the fourth power. Furthermore, the integral of functions of third degree, whose third derivative is constant, is exactly represented by our formula, for all the terms of the correction are then zero.[7]

A formula of about the same accuracy as Simpson's rule, which is commonly known as *Newton's formula,* is obtained in the following way. We write the Euler formula for $\kappa = 3\mu$:

$$J_1 = h\sum_0^{3\mu} y_\lambda - h\frac{y_0 + y_{3\mu}}{2} - \frac{h^2}{12}[f'(x_{3\mu}) - f'(x_0)]$$
$$+ \frac{h^4}{720}[f'''(x_{3\mu}) - f'''(x_0)] - \frac{h^6}{30240}[f^{(5)}(x_{3\mu}) - f^{(5)}(x_0)] \cdots$$

and then write the Euler formula for the same region, but for subintervals of triple width:

$$J_2 = 3h\sum_0^{\mu} y_{3\lambda} - 3h\frac{y_0 + y_{3\mu}}{2} - \frac{3h^2}{4}[f'(x_{3\mu}) - f'(x_0)]$$
$$+ \frac{9h^4}{80}[f'''(x_{3\mu}) - f'''(x_0)] - \frac{27h^6}{1120}[f^{(5)}(x_{3\mu}) - f^{(5)}(x_0)] \cdots .$$

Now the terms with the first derivative can be made to vanish, if we form

$$(25)\qquad J_N = \frac{9J_1 - J_3}{8} = \frac{3h}{8}\sum_0^{\mu}(y_{3\lambda} + 3y_{3\lambda+1} + 3y_{3\lambda+2} + y_{3\lambda+3}) + K_N\ ,$$

where the correction is

$$(26)\ K_n = -\frac{1}{80}h^4(f'''(x_{3\mu}) - f'''(x_0)) + \frac{1}{336}h^6(f^{(5)}(x_{3\mu}) - f^{(5)}(x_0))\ \cdots\ .$$

If the number of subintervals is divisible by 6, we can go a step further. By combination of the Simpson and Newton formulas, still another term of the correction can be made to vanish. If we set $\nu = 3\rho$ and $\mu = 2\rho$, we get the so-called *Weddle* formula:[8]

$$(27)\qquad \begin{aligned} J_W = \frac{18J_S - 8J_N}{10} = \frac{3h}{10}\sum_0^{\rho}(y_{6\lambda} + 5y_{6\lambda+1} + y_{6\lambda+2} + 6y_{6\lambda+3} \\ + y_{6\lambda+4} + 5y_{6\lambda+5} + y_{6\lambda+6}) + K_W\ , \end{aligned}$$

where now

$$(28)\qquad K_W = -\frac{h^6}{840}(f^{(5)}(x_{6\rho}) - f^{(5)}(x_0))\ \cdots\ .$$

There is a large number of similar formulas, but these shall be sufficient.

5. *Example*: The integral $\int_0^{\pi/2} \sin x\, dx$ is to be calculated by the above formula from the values $\sin 0° = 0.000000$, $\sin 15° = 0.258819$, $\sin 30° = 0.500000$, $\sin 45° = 0.707107$, $\sin 60° = 0.866025$, $\sin 75° = 0.965926$, $\sin 90° = 1.000000$.

From these, we obtain

$$y_T = 3.797877 \cdot \frac{\pi}{12} = 0.994282,$$

$$y_L = 1.931852 \cdot \frac{\pi}{6} = 1.011515,$$

$$y_S = 11.459457 \cdot \frac{\pi}{36} = 1.000026,$$

$$y_N = 10.186524 \cdot \frac{3\pi}{96} = 1.000059,$$

$$y_W = 12.732390 \cdot \frac{3\pi}{120} = 1.000000.$$

The magnitude of the error corresponds roughly to the correction term given above.

If we have function values which contain errors, we must not use formulas in which the values of the function are multiplied by very different weights. For then the corresponding errors are multiplied by these weights also. In such cases the use of the trapezoid rule or MacLaurin's formula is much more to be recommended.

6. As a *second example,*[9] we show how, under certain circumstances, integrals in which a parameter appears can be evaluated by this formula. The oscillation period of the simple pendulum with an amplitude α is given by

$$\begin{aligned} T &= 2\left(\frac{l}{g}\right)^{1/2} \int_0^{\pi/2} \frac{d\psi}{(1 - \sin^2(\alpha/2)\sin^2\psi)^{1/2}} \\ &= \pi\left(\frac{l}{g}\right)^{1/2}\left(1 + \left(\frac{1}{2}\right)^2 \sin^2\frac{\alpha}{2} + \left(\frac{1\cdot 3}{2\cdot 4}\right)^2 \sin^4\frac{\alpha}{2} \right. \\ &\qquad \left. + \left(\frac{1\cdot 3\cdot 5}{2\cdot 4\cdot 6}\right)^2 \sin^6\frac{\alpha}{2} \cdots\right) \end{aligned} \tag{29}$$

where a series expansion replaces the integral. If now we use the trapezoid rule for one interval width on the above integral, we obtain

$$\begin{aligned} J_T &= \frac{\pi}{2}\left(\frac{l}{g}\right)^{1/2}\left(1 + \frac{1}{(1 - \sin^2\alpha/2)^{1/2}}\right) = \frac{\pi}{2}\left(\frac{l}{g}\right)^{1/2}\left(\frac{1 + \cos\alpha/2}{\cos\alpha/2}\right) \\ &= \pi\left(\frac{l}{g}\right)^{1/2}\left(1 + \left(\frac{1}{2}\right)^2 \sin^2\frac{\alpha}{2} + \frac{1}{2}\cdot\frac{1\cdot 3}{2\cdot 4}\sin^4\frac{\alpha}{2} \right. \\ &\qquad \left. + \frac{1}{2}\cdot\frac{1\cdot 3\cdot 5}{2\cdot 4\cdot 6}\sin^6\frac{\alpha}{2}\cdots\right). \end{aligned} \tag{30}$$

The correction for this formula is then

$$K_T = -\pi\left(\frac{l}{g}\right)^{1/2}\left[\frac{3}{64}\sin^4\frac{\alpha}{2} + \frac{15}{256}\sin^6\frac{\alpha}{2} + \frac{35\cdot 29}{128^2}\sin^8\frac{\alpha}{2}\cdots\right]. \tag{31}$$

If we approximate by MacLaurin's formula, we get

$$\begin{aligned} J_L &= \pi\left(\frac{l}{g}\right)^{1/2} \frac{1}{(1 - (1/2)\sin^2\alpha/2)^{1/2}} \\ &= \pi\left(\frac{l}{g}\right)^{1/2}\left(1 + \left(\frac{1}{2}\right)^2 \sin^2\frac{\alpha}{2} + \frac{1\cdot 3}{2\cdot 4}\cdot\frac{1}{4}\sin^4\frac{\alpha}{2} \right. \\ &\qquad \left. + \frac{1\cdot 3\cdot 5\cdot 1}{2\cdot 4\cdot 6\cdot 8}\sin^6\frac{\alpha}{2}\cdots\right). \end{aligned} \tag{32}$$

The corresponding correction is

$$(33)\quad K_L = \pi\left(\frac{l}{g}\right)^{1/2}\left(\frac{3}{64}\sin^4\frac{\alpha}{2} + \frac{15}{256}\sin^6\frac{\alpha}{2} + \frac{35\cdot 27}{128\cdot 128}\sin^8\frac{\alpha}{2}\cdots\right).$$

Here Simpson's rule gives essentially no better approximation, as can be seen immediately. On the other hand, the first terms of the correction series cancel if we take the arithmetic mean of the two approximation formulas above. We then have

$$(34)\qquad J = \frac{\pi}{2}\left(\frac{l}{g}\right)^{1/2}\left(\frac{1}{2} + \frac{1}{(1-(1/2)\sin^2\alpha/2)^{1/2}} + \frac{1}{2\cos\alpha/2}\right),$$

with the correction

$$(35)\qquad K = -\pi\left(\frac{l}{g}\right)^{1/2}\left(\frac{35}{128^2}\sin^8\alpha\cdots\right).$$

NOTES

1. Euler, *Comm. Acad. Soc. Imp. Petrop.*, 6 (1738).
2. MacLaurin, *Treatise of fluxions* (1742).
3. Bernoulli. *Ars conjectandi* (Basel, 1713). Norlund, *Differenzenrechnung* (Berlin, 1924), p. 457.
4. Jacob Bernoulli, *Ars conjectandi* (1713).
5. Lubbock, *Camb. Phil. Trans.* 3 (1829), p. 323.
6. Simpson, *Math. Dissertation* (1743), pp. 109-119.
7. If the number of intervals is not divisible by two, that is, if there is an extra interval, we assume that it is the first. It is then customary to use for the calculation of the integral in this interval the formula

$$(23)\qquad \bar{y}_S = \frac{h}{12}(5y_0 + 8y_1 - y_2),$$

which is known as the special Simpson's rule. To investigate its accuracy, we develop y in a series about the point x_0:

$$y = f(x_0) + f'(x_0)h + f''(x_0)\frac{h^2}{2} + f'''(x_0)\frac{h^3}{6} + f^{(4)}(x_0)\frac{h^4}{24} + \cdots.$$

The correct value of the integral for this interval is then

$$\int_{x_0}^{x_0+h} y\,dx = f(x_0)h + f'(x_0)\frac{h^2}{2} + f''(x_0)\frac{h^3}{6}$$

$$+ f'''(x_0)\frac{h^4}{24} + f^{(4)}(x_0)\frac{h^5}{120}\cdots.$$

In addition, we find

$$5hy_0 = 5hf(x_0)$$

$$8hy_1 = 8hf(x_0) + 8h^2f'(x_0) + 4h^3f''(x_0) + \frac{4}{3}f'''(x_0)$$

$$+ \frac{1}{3}h^5f^{(4)}(x_0) - hy_2 = -hf(x_0) - 2h^2f'(x_0) - 2h^3f''(x_0)$$

$$- \frac{4}{3}h^4f'''(x_0) - \frac{2}{3}h^5f^{(4)}(x_0).$$

By collection of terms and division by 12, we get

$$\bar{y}_S = hf(x_0) + \frac{h^2}{2}f'(x_0) + \frac{h^3}{6}f''(x_0) - \frac{h^5}{36}f^{(4)}(x_0).$$

The correction is then

$$\bar{\kappa}_S = \frac{1}{24}f'''(x_0)h^4 + \frac{13}{360}f^{(4)}(x_0)h^5. \tag{24}$$

The term in which the third derivative appears is then in error. This term is still given correctly by Simpson's rule. The formula is occasionally used for integration of differential equations (33.5) because of the small factors with which the third ordinate is multiplied.

8. Weddle, *Cambridge math. Journal* (1854), p. 79.
9. Poukka, *Zeits. f. angew. Math. u. Mech. V* (1925), p. 521.

16. Mean Value Methods.

1. In addition to the methods given in Art. 12, which use differences for the calculation of definite integrals, a series of formulas have been produced which seek to approximate the value of the integral by a mean value of the particular function values of the integration interval, multiplied by a corresponding weight. If we consider the function plotted as a curve, the *surface area* under this curve is approximated *by the area under a curve of* $(n - 1)$*st degree which passes through the end points of the ordinates entering into the calculation.* For the derivation of this formula, we start from the simplest Lagrange interpolation formula given in 8(21):

$$f(x) = \sum_1^n \frac{f(x_r)\varphi(x)}{(x - x_r)\varphi'(x_r)} + [xx_1 \cdots x_n]\varphi(x), \tag{1}$$

where $\varphi(x) = (x - x_1)(x - x_2) \cdots (x - x_n)$ is an integral function of the nth degree with roots $x_1, x_2, \cdots, x_n$. If this expression is integrated between the limits a and b, there results

$$\int_a^b f(x)\,dx = \sum_1^n B_r f(x_r) + \int_a^b [xx_1 \cdots x_n]\varphi(x)\,dx, \tag{2}$$

where

$$B_r = \int_a^b \frac{\varphi(x)}{(x - x_r)\varphi'(x_r)}\,dx \tag{3}$$

is written for the weights with which the individual function values are to be multiplied. If we set $f(x) = 1$, we obtain for the sum of these weights, independent of $f(x)$,

$$(b - a) = \sum_1^n B_r\,. \tag{4}$$

Except for the remainder term, then, the arithmetic mean $\int_a^b f(x)\,dx/(b - a)$ is expressed as the mean of the separate function values, multiplied by the corresponding weights, i.e., by

$$\frac{1}{b - a}\int_a^b f(x)\,dx = \frac{\sum_1^n B_r f(x_r)}{\sum_1^n B_r}. \tag{5}$$

For the following considerations, it is practical to change the limits of integration to $\pm h$ by the substitution

$$x = \frac{b + a}{2} + \frac{b - a}{2h}\cdot t, \qquad dx = \frac{b - a}{2h}\,dt, \tag{6}$$

so that the integral becomes

$$\int_a^b f(x)\,dx = \frac{b - a}{2h}\int_{-h}^{+h} f(t)\,dt. \tag{7}$$

We shall now consider only the latter integral, and set

$$\int_{-h}^{+h} f(t)\,dt = \sum_1^n A_r f(t_r) + \int_{-h}^{+h} [t, t_1, t_2 \cdots t_n]\varphi(t)\,dt. \tag{8}$$

2. The function values to be used for the approximation can be chosen with complete freedom. The weights with which the various function values are to be multiplied are determined by this choice. In general, *the function values will be chosen symmetrically about the middle of the interval.* In this case we have, for an even number of ordinates, $n = 2m$,

$$\varphi(t) = (t^2 - t_1^2)(t^2 - t_2^2) \cdots (t^2 - t_m^2), \tag{9}$$

and for an odd number, $n = 2m + 1$,

$$\varphi(t) = t(t^2 - t_1^2)(t^2 - t_2^2) \cdots (t^2 - t_m^2). \tag{10}$$

Then the weights belonging to the symmetric ordinates A_κ and $A_{-\kappa}$ are equal; e.g., for the even number, $n = 2m$,

$$A_\kappa = \int_{-h}^{+h} \frac{(t^2 - t_1^2)(t^2 - t_2^2) \cdots (t^2 - t_m^2)}{(t - t_\kappa)\cdot 2t_\kappa(t_\kappa^2 - t_1^2) \cdots (t_\kappa^2 - t_m^2)}\, dt, \tag{11}$$

where the factor $(t_\kappa^2 - t_\kappa^2)$ does not appear in the denominator. If t_κ is replaced here by $-t_\kappa$, the parentheses containing the squared terms are not changed. In the denominator there appears $-(t + t_\kappa)2t_\kappa$ instead of $(t - t_\kappa)2t_\kappa$. If further the integration variable is replaced by $-t$, then the parentheses containing the squares are again unchanged and we have

$$\begin{aligned} A_{-\kappa} &= \int_{+h}^{-h} \frac{(t^2 - t_1^2)(t^2 - t_2^2) \cdots (t^2 - t_m^2)}{(t - t_\kappa)2t_\kappa(t_\kappa^2 - t_1^2) \cdots (t_\kappa^2 - t_m^2)} \cdot (-dt) \\ &= \int_{-h}^{+h} \frac{(t^2 - t_1^2)(t^2 - t_2^2) \cdots (t^2 - t_m^2)}{(t - t_\kappa)2t_\kappa(t_\kappa^2 - t_1^2) \cdots (t_\kappa^2 - t_m^2)}\, dt = A_\kappa\,. \end{aligned} \tag{12}$$

For the case of an odd number of ordinates, $n = 2m + 1$ only the factor t/t_κ is added under the integral in the integrand above. This factor does not change sign with the substitution of $-t_\kappa$ for t_κ, and $-t$ for t. In general, therefore, when the values of the function lie symmetrically about the middle, the *weights associated with these symmetric function values are equal.*

3. The oldest choice of function values, as is found in Newton[1] and Cotes[2], is one in which the function values chosen are equally spaced, making use of the end ordinates of the interval. Function values are then to be chosen for the following abscissas for even $n = 2m$,

$$\begin{aligned} &-h;\ -\frac{n-3}{n-1}h;\ -\frac{n-5}{n-1}h;\ \cdots\ -\frac{h}{n-1},\ +\frac{h}{n-1}, \\ &\qquad +\frac{3h}{n-1} + \cdots + \frac{n-3}{n-1}h,\ +h \end{aligned} \tag{12a}$$

and for odd $n = 2m + 1$,

$$\begin{aligned} &-h;\ -\frac{n-3}{n-1}h;\ -\frac{n-5}{n-1}h\ \cdots,\ -\frac{2h}{n-1},\ 0,\ +\frac{2h}{n-1}, \\ &\qquad +\frac{4h}{n-1} + \cdots + \frac{n-3}{n-1}h,\ +h. \end{aligned} \tag{12b}$$

The coefficients are given by the integral above, in which these values are substituted. For even $n = 2m$, we have

$$A_\kappa = \frac{h}{n-1}\int_{-(n-1)}^{+(n-1)} \frac{(u^2-1^2)(u^2-3^2)\cdots(u^2-[n-1]^2)\cdot du}{(u-[2\kappa-1])(4\kappa-2)(2\kappa-2)\cdot 2\kappa(2\kappa-4)(2\kappa+2)\cdots 2(4\kappa-4)(-2)\cdot 4\kappa\cdots(2\kappa-n)(2\kappa+n-2)} \tag{13}$$

$$= \frac{h(-1)^{(n/2)-\kappa}}{(n-1)2^{n-1}((n/2)-\kappa)!((n/2)+\kappa-1)!}\int_{-(n-1)}^{+(n-1)} \frac{(u^2-1^2)(u^2-3^2)\cdots(u^2-[n-1]^2)du}{u-(2\kappa-1)},$$

if we introduce u as a new variable, given by $t = hu/(n-1)$,

Example: For $n = 4$ and $\kappa = 1$, we get

$$A_{\pm 1} = \frac{-h}{3\cdot 8\cdot 1!\cdot 2!}\int_{-3}^{+3} (u+1)(u^2-9)\,du = -\frac{h}{48}\left[\frac{u^4}{4} + \frac{u^3}{3} - \frac{9u^2}{2} - 9u\right]_{-3}^{+3} = 2h\cdot\frac{3}{8}, \tag{13a}$$

and for $n = 4$, $\kappa = 2$,

$$A_{\pm 2} = \frac{h}{3\cdot 8\cdot 0!\cdot 3!}\int_{-3}^{+3} (u^2-1)(u+3)\,du = \frac{h}{3\cdot 48}\left[\frac{u^4}{4} + u^3 - \frac{u^2}{2} - 3u\right]_{-3}^{+3} = 2h\cdot\frac{1}{8}. \tag{13b}$$

We then get

$$J = \frac{2h}{8}(y_{-1} + 3y_{-1/3} + 3y_{1/3} + y_{+1}) + R.$$

This is the formula designated as Newton's formula in section 4 of the preceding article, applied to an interval of width $2h$, instead of $3h$.

For an odd number of ordinates, we get the coefficients, after the same substitution,

$$A_\kappa = \frac{h}{n-1}\int_{-(n-1)}^{+(n-1)} \frac{u(u^2-2^2)(u^2-4^2)\cdots(u^2-[n-1]^2)du}{(u-2\kappa)\cdot 4\kappa\cdot 2\kappa(2\kappa-2)(2\kappa+2)\cdots(+2)(4\kappa-2)(-2)(4\kappa+2)\cdots(2\kappa-n+1)(2\kappa+n-1)}$$

$$= \frac{h(-1)^{[(n-1)/2]-\kappa}}{(n-1)2^{n-1}(\kappa+(n-1)/2)!((n-1)/2-\kappa)!}\int_{-(n-1)}^{+(n-1)} \frac{u(u^2-2^2)\cdots(u^2-[n-1]^2)}{u-2\kappa}\,du \tag{14}$$

$$= \frac{h(-1)^{m-\kappa}}{m(m+\kappa)!(m-\kappa)!}\int_{-m}^{+m} \frac{v(v^2-1^2)\cdots(v^2-m^2)}{v-\kappa}\,dv,$$

where $u = 2v$.

Example: For $n = 3$, we have

$$A_0 = \frac{-h}{1\cdot 1!\cdot 1!}\int_{-1}^{+1}(v^2 - 1)\,dv = -h\left[\frac{v^2}{3} - v\right]_{-1}^{+1} = \frac{4}{3}h,$$

(14a)

$$A_{\pm 1} = \frac{h}{1\cdot 2!\cdot 0!}\int_{-1}^{+1} v(v+1)\,dv = \frac{h}{2}\left[\frac{v^3}{3} + \frac{v^2}{2}\right]_{-1}^{+1} = \frac{1}{3}h$$

and therefore

$$y = \frac{2h}{6}[y_{-1} + 4y_0 + y_{+1}] + R,$$

i.e., the Kepler rule which was derived in other ways in 12(12) and 15(21). By suitable calculation, the following numerical values are obtained:

$n = 2 \; A_{\pm 1} = h,$

$n = 3 \quad A_0 = \frac{4h}{3} \quad A_{\pm 1} = \frac{h}{3},$

$n = 4 \; A_{\pm 1} = \frac{3h}{4} \quad A_{\pm 2} = \frac{h}{4}$

$n = 5 \quad A_0 = \frac{4}{15}h \quad A_{\pm 1} = \frac{32}{45}h \quad A_{\pm 2} = \frac{7}{45}h,$

$n = 6 \; A_{\pm 1} = \frac{25}{72}h \quad A_{\pm 2} = \frac{25}{48}h \quad A_{\pm 3} = \frac{19}{144}h,$

$n = 7 \quad A_0 = \frac{68}{105}h \; A_{\pm 1} = \frac{9}{140}h \; A_{\pm 2} = \frac{18}{35}h \quad A_{\pm 3} = \frac{41}{420}h,$ etc.[3]

4. The *remainder term of the formula* contains an nth divided difference. Now it can be shown that in general the formulas with an odd number of ordinates are the more advantageous if the value of the approximation is estimated by the number of the terms of the Taylor expansion of the integral which will give the approximation value exactly. This has already been done in 15.4 for Simpson's and Newton's formulas for the case of a single interval. We limit ourselves here to this case, where $n = 2m + 1$. According to 16(8) and 16(10) then, the remainder term has the value

$$R = \int_{-h}^{+h}[t, t_0, t_1, -t_1, \cdots, t_m, -t_m]t(t^2 - t_1^2)\cdots(t^2 - t_m^2)\,dt,$$

or, if we use the values of t_k given above, introduce a new variable v, given by $2hv = (n - 1)\,t = 2mt$, and consider that the expression in brackets then becomes a function of v,

$$(15)\quad R = \left(\frac{h}{m}\right)^{2m+2} \int_{-m}^{+m} [t, t_0, t_1, -t_1, \cdots, t_m, -t_m] v(v^2 - 1)(v^2 - 2^2) \cdots (v^2 - m^2)\, dv.$$

We transform this according to Steffensen[4]: For abbreviation, we set $P(v) = v(v^2 - 1)(v^2 - 2) \cdots (v^2 - m^2)$; also, $Q(v) = \int_{-m}^{v} P(v)\, dv$. Obviously, $Q(-m) = 0$, but also $Q(+m) = 0$, since $P(v)$ is an odd function: $P(-v) = -P(v)$. Now, to show that $Q(v)$ is never zero in the interval from $-m$ to $+m$, we examine

$$I_r = \int_r^{r+1} P(v)\, dv,$$

where r and $r + 1$ are consecutive roots of $P(v)$. It can easily be seen that the relation

$$(16)\quad I_{-r} = \int_{-r}^{-r+1} P(v)\, dv = \int_{r-1}^{r} P(-v)\, dv = -\int_{r-1}^{r} P(v)\, dv = -I_{r-1}$$

exists. If we now go back to the value of $P(v)$, we have

$$I_{r-1} = \int_{r-1}^{r} (v + m)(v + m - 1) \cdots (v - m)\, dv$$

$$(16a)\qquad = \int_r^{r+1} (v + m - 1)(v + m - 2) \cdots (v - m - 1)\, dv$$

$$= \int_r^{r+1} \frac{v - m - 1}{v + m} P(v)\, dv.$$

Now, since $P(v)$ does not vanish in the observed interval, the mean value theorem can be applied and we get

$$(16b)\quad I_{r-1} = \frac{\bar{v} - m - 1}{\bar{v} + m} \int_r^{r+1} P(v)\, dv = \frac{\bar{v} - m - 1}{\bar{v} + m} I_r \quad (r < \bar{v} < r + 1).$$

Since $\bar{v} < m$, $\bar{v} > 0$, then $|I_r| > |I_{r-1}|$. The two have opposite signs because of the sign of $P(v)$. Upon consideration of (16), it follows, if $I_{-m} = 0$, that

$$I_{-m} > -I_{-m+1} > I_{-m+2} \cdots > \pm I_{-2} > \mp I_{-1}$$

(the sign of $P(v)$ can be chosen to correspond), i.e., the function $Q(v)$ cannot be zero for negative v, nor for $v = 0$. Since the function is sym-

metric with respect to the y axis, $Q(v)$ does not change sign between $-m$ and $+m$. If we integrate (15) by parts, and remember that $Q(-m) = Q(+m) = 0$, that $d/dv = (h/m)d/dt$, and also recall 11(2), we get

$$R = -\left(\frac{h}{m}\right)^{2m+3} \int_{-m}^{+m} Q(v)[t, t, t_0, t_1, -t_1 \cdots t_m, -t_m]\, dv.$$

Since, as was shown above, $Q(v)$ does not change its sign in the interval, we can apply the mean value theorem of the integral calculus. Therefore we have

$$R = -\left(\frac{h}{m}\right)^{2m+3} [\bar{t}, \bar{t}, t_0, t_1, -t_1, \cdots, t_m, -t_m] \int_{-m}^{+m} Q(v)\, dv,$$

$$(-t_m < \bar{t} < t_m).$$

If we integrate the latter integral by parts once again, we get

$$\int_{-m}^{+m} Q(v)\, dv = [vQ(v)]_{-m}^{+m} - \int_{-m}^{+m} vQ'(v)\, dv = -\int_{-m}^{+m} vP(v)\, dv.$$

Therefore, by 8(24), if $f^{(2m+2)}$ exists and is continuous in the interval,

$$R = \frac{2f^{(2m+2)}(\tau)}{(2m+2)!}\left(\frac{h}{m}\right)^{2m+3} \int_0^m v^2(v^2 - 1)(v^2 - 2^2) \cdots (v^2 - m^2)\, dv,$$

$$-t_m < \tau < t_m,$$

where the integral is always negative so that the remainder term has the opposite sign to the $(2m + 2)$nd derivative at the point in question.

For example, for the case $n = 3$, for which the weights have been calculated above, there results

$$R_3 = 2h^3 \frac{f^{(4)}(\tau)}{4!} \int_0^1 v^2(v^2 - 1)\, dv = -\frac{4}{15}\frac{h^5 f^{(4)}(\tau)}{4!}.$$

In the cases $n = 5$ and $n = 7$, for which the weights are also given above, the remainders are

$$R_5 = -\frac{1}{21} h^7 \frac{f^{(6)}(\tau)}{6!}; \qquad R_7 = -\frac{16}{1215} h^9 \frac{f^{(8)}(\tau)}{8!}.$$

5. With a second group of formulas, the entire interval is divided into equal subintervals. But the end ordinates of the main interval and the bounding ordinates of the various subintervals are not used for the formation of the mean values. On the contrary, we use the middle ordinate of each subinterval. These formulas are usually called *MacLaurin's formulas*, although only the formula derived in 15(19), which coincides ($\kappa = 1$) with the first formula derived here, is due to him.

In the case of an even number of function values, the ordinates for the abscissa values are

$$-\frac{n-1}{n}, \frac{n-3}{n}, \cdots, -\frac{1}{n}, +\frac{1}{n}, +\frac{3}{n}, \cdots, \frac{n-3}{n}, \frac{n-1}{n},$$

and for an odd number of ordinates,

$$-\frac{n-1}{n}, -\frac{n-3}{n}, \cdots, -\frac{2}{n}, 0, +\frac{2}{n}, \cdots, \frac{n-3}{n}, \frac{n-1}{n}.$$

If we substitute the values for t_κ, introduce u as a new variable by the equation $hu = nt$, and set $n = 2m$, the coefficients in the first case become

$$A_\kappa = \frac{(-1)^{(n/2)-\kappa}\cdot h}{n\cdot 2^{n-1}(n/2+\kappa-1)!(n/2-\kappa)!}$$

$$\int_{-n}^{+n} \frac{(u^2-1^2)(u^2-3^2)\cdots(u^2-(n-1)^2)}{u-(2\kappa-1)}\,dy$$

(17)

$$= \frac{(-1)^{m-\kappa}\cdot h}{2^{2m}\cdot m(m+\kappa-1)!(m-\kappa)!}$$

$$\int_{-2m}^{+2m} \frac{(u^2-1^2)(u^2-3^2)\cdots(u^2-(2m-1)^2)}{u-(2\kappa-1)}\,du.$$

For example, for $n = 4$, this gives

$$A_{\pm 1} = \frac{-h}{2^4\cdot 2\cdot 2!\cdot 1!}\int_{-4}^{+4}(u+1)(u^2-9)\,du$$

$$= -\frac{h}{64}\left[\frac{u^4}{4}+\frac{u^3}{3}-\frac{9}{2}u^2-9u\right]_{-4}^{+4} = \frac{11h}{24},$$

(17a)

$$A_{\pm 2} = \frac{h}{2^4\cdot 2\cdot 3!\cdot 0!}\int_{-4}^{+4}(u^2-1)(u+3)\,du$$

$$= \frac{h}{192}\left[\frac{u^4}{4}+u^3-\frac{u^2}{2}-3u\right]_{-4}^{+4} = \frac{13h}{24}.$$

The second coefficient could also have been calculated on the basis that the sum of the weights must be $2h$. This fact is better used for checking the calculation. The formula is then

(17b) $$y = \frac{2h}{48}\left[13(y_{-(3/4)} + y_{+(3/4)}) + 11(y_{-(1/4)} + y_{+(1/4)})\right] + R.$$

With these formulas also, the best approximation formula is given for an odd number of ordinates; that means an odd number of subintervals also. In the case $n = 2m + 1$ if we introduce the abscissa values given above in the integral, and set $uh = nt$ and $u = 2v$, then the relation becomes

(18) $$\begin{aligned} A_\kappa &= \frac{h(-1)^{m-\kappa}}{n2^{n-1}(m+\kappa)!(m-\kappa)!} \\ &\qquad \int_{-(2m+1)}^{+(2m+1)} \frac{u(u^2-2^2)(u^2-4^2)\cdots(u^2-(n-1)^2)}{u-2\kappa}\,du \\ &= \frac{2h(-1)^{m-\kappa}}{(2m+1)(m+\kappa)!(m-\kappa)!} \\ &\qquad \int_{(-m+(1/2))}^{+m+(1/2)} \frac{v(v^2-1^2)(v^2-2^2)\cdots(v^2-m^2)}{v-\kappa}\,dv. \end{aligned}$$

For example, for $n = 3$,

(18a) $$\begin{aligned} A_0 &= \frac{-2h}{3\cdot 1!\cdot 1!}\int_{-3/2}^{+3/2}(v^2-1)\,dv = -\frac{2h}{3}\left[\frac{v^3}{3} - v\right]_{-3/2}^{+3/2} = \frac{h}{2}, \\ A_{\pm 1} &= \frac{2h}{3\cdot 2!\cdot 0!}\int_{-3/2}^{+3/2} v(v+1)\,dv = \frac{2h}{6}\left[\frac{v^3}{3} + \frac{v^2}{2}\right]_{-3/2}^{+3/2} = \frac{3h}{4}. \end{aligned}$$

Then we have

$$y = \frac{2h}{8}\left[2y_0 + 3(y_{+3/2} + y_{-3/2})\right] + R.$$

The coefficients of the other formulas may be calculated in the same way. The following values are obtained:

$n = 1$: $A_0 = 2h$,

$n = 2$: $A_{+1} = h$,

$n = 3$: $A_0 = \frac{h}{2}$, $\qquad A_{\pm 1} = \frac{3}{4}h$,

$n = 4$: $A_{\pm 1} = \frac{11}{24}h$, $\qquad A_{\pm 2} = \frac{13}{24}h$,

$n = 5$: $A_0 = \frac{67}{96}h$, $\qquad A_{\pm 1} = \frac{25}{144}h$, $\quad A_{\pm 2} = \frac{275}{576}h$,

$$n = 6: A_{\pm 1} = \frac{127}{320}h, \qquad A_{\pm 2} = \frac{139}{640}h, \quad A_{\pm 3} = \frac{247}{640}h,$$

$$n = 7: \quad A_0 = -\frac{6257}{17280}h, \; A_{\pm 1} = \frac{6223}{7680}h, \; A_{\pm 2} = \frac{49}{3840}h,$$

$$A_{\pm 3} = \frac{4949}{13824}h.$$

Negative weights appear in the last formula above.

6. As was mentioned previously, the formulas with an odd number of ordinates, and also with an odd number of subintervals, are the more advantageous. If the function is sufficiently differentiable, and if the $(2m + 2)$nd derivative is continuous in the closed interval of integration, then the remainder term can be put in the form due to Walter,[5] in the case $n = 2m + 1$:

$$(19) \quad R_n = 2\left(\frac{h}{m}\right)^{2m+3} \frac{f^{(2m+2)}(\tau)}{(2m+2)!} \int_0^{m+1/2} v^2(v^2 - 1^2) \cdots (v^2 - m^2)\, dv$$

where $-h \leqq \tau \leqq +h$. It can be shown that this integral is always positive, so that the correction has the same sign as the $(n + 1)$st derivative.

In the case $n = 3$ calculated above, for example, we have

$$(19a) \quad \begin{aligned} R_3 &= 2h^5 \frac{f^{(4)}(\tau)}{4!} \int_0^{3/2} v^2(v^2 - 1)\, dv \\ &= \frac{1}{12} h^5 f^{(4)}(\tau) \left[\frac{v^5}{5} - \frac{v^3}{3}\right]_0^{3/2} = \frac{21}{640} h^5 f^{(4)}(\tau). \end{aligned}$$

In the same way we get

$$(19b) \quad R_5 = \frac{5575}{193536}\left(\frac{h}{2}\right)^7 f^{(6)}(\tau), \qquad R_7 = \frac{1718381}{66355200}\left(\frac{h}{3}\right)^9 f^{(8)}(\tau).$$

7. In the formulas derived so far, the individual function values are multiplied by various weights. If now we consider function values determined empirically, then these themselves have some errors. It is not advantageous in mean value methods to multiply the errors with different weights. The error of the result is a minimum if the factors of the individual errors are all equal. Such formulas were first derived by *Tschebyscheff*[6] and therefore have been named after him.

In this case the coefficients A_κ are prescribed and the integrals (11) give n equations for these coefficients. From these the n unknown abscissas

can be calculated for which the function values are to be chosen. These equations are not independent of each other. There exists among them the relation $\sum_1^n A_\kappa = 2h$. The A_κ could be given arbitrary values which satisfy this equation. If the values of the function are further assumed to be symmetrical, then the abscissa of a function value $x = 0$ is determined for odd $n = 2m + 1$. Now $2m$ values are to be calculated from the $2m + 1$ equations among which a linear relation exists. Therefore only $2m$ of these are independent. In this case the $2m$ equations are reduced to m because the coefficients belonging to the symmetric function values have the same equation. The integrands coincide for odd n in the numerator, that is, except for the factor $(t - t_\kappa)$ or $(t + t_\kappa)$. The numerator is $t(t^2 - t_1^2) \cdots (t^2 - t_m^2)$, except for these factors, and therefore contains only odd powers. But only even powers make a contribution to the integral in integration between the limits $-h$ and h, for the upper and lower limit values of the integral cancel each other in the case of odd powers in the integrand. Therefore, only the part of the integrand multiplied by t, and not that multiplied by $\pm t_\kappa$, makes a contribution to A_κ. These parts are equal for A_κ and $A_{-\kappa}$. Consequently the same equations are obtained. We limit ourselves here to the equal values $A_\kappa = 2h/n$.

Example: $n = 5$. The equations become

$$A_0 = \frac{2h}{5} = \int_{-h}^{+h} \frac{(t^2 - t_1^2)(t^2 - t_2^2)}{t_1^2 \cdot t_2^2}\, dt$$

$$= \frac{1}{t_1^2 t_2^2}\left(\frac{2h^5}{5} - (t_1^2 + t_2^2)\frac{2h^3}{3} + 2t_1^2 t_2^2 h\right).$$

(19c)

$$A_{\pm 1} = \frac{2h}{5} = \int_{-h}^{+h} \frac{t(t + t_1)(t^2 - t_2^2)}{2t_1^2(t_1^2 - t_2^2)}\, dt$$

$$= \frac{1}{2t_1^2(t_1^2 - t_2^2)}\left(\frac{2h^5}{5} - \frac{2h^3}{3}\, t_2^2\right),$$

$$A_{\pm 2} = \frac{2h}{5} = \int_{-h}^{+h} \frac{t(t + t_2)(t^2 - t_1^2)}{2t_2^2(t_2^2 - t_1^2)}\, dt$$

$$= \frac{1}{2t_2^2(t_2^2 - t_1^2)}\left(\frac{2h^5}{5} - \frac{2h^3}{3}\, t_1^2\right).$$

Since the equations depend on each other, we calculate the values of $\pm t_1$ and $\pm t_2$ from the last two. The equations are

$$\frac{4}{5}t_1^2(t_1^2 - t_2^2) = \frac{2}{5}h^4 - \frac{2}{3}h^2t_2^2,$$

$$\frac{4}{5}t_2^2(t_2^2 - t_1^2) = \frac{2}{5}h^4 - \frac{2}{3}h^2t_1^2.$$

From these we obtain by subtraction

$$t_2^2 + t_1^2 = \frac{5}{6}h^2,$$

and by substitution in one of the equations:

$$t_1^4 - \frac{5}{6}h^2t_1^2 = -\frac{7}{72}h^4.$$

From this it follows that

$$t_1^2 = \frac{5 + (11)^{1/2}}{12}h^2; \qquad t_2^2 = \frac{5 - (11)^{1/2}}{12}h^2.$$

We can be sure that these values also satisfy the first equation.

In the case of an even number of function values, the weight $A_0 = 0$ can be given to the function value for $x = 0$, and the preceding equations may then be used. Then, for symmetric function values, m equations are again given for the determination of the m squares of the abscissas, at which the function values are to be chosen.

Example: For $n = 4$, we have the same equations as for $n = 5$, except that we set $A_0 = 0$, $A_{\pm 1} = A_{\pm 2} = h/2$. The three equations then become

$$\frac{h^4}{5} - (t_1^2 + t_2^2)\frac{h^2}{3} + t_1^2t_2^2 = 0,$$

$$\frac{2h^4}{5} - \frac{2h^2}{3}t_1^2 = t_2^2(t_2^2 - t_1^2),$$

$$\frac{2h^4}{5} - \frac{2h^2}{3}t_2^2 = t_1^2(t_1^2 - t_2^2).$$

From the last two we obtain the equation

$$t^4 - \frac{2}{3}h^2t^2 = -\frac{h^4}{45},$$

from which it follows that

$$t_1^2 = \frac{5 + 2(5)^{1/2}}{15} h^2, \qquad t_2^2 = \frac{5 - 2(5)^{1/2}}{15} h^2.$$

It can easily be demonstrated that the first equation is satisfied. Therefore about the same approximation is obtained with $2m$ function values as with $2m + 1$, since the approximation is made with a curve of $2m$th degree in each case. The formulas with an even number of ordinates are therefore the more practical.

The remainder term has the form

$$(22) \qquad R_n = \int_{-h}^{+h} [t, 0, t_1, -t_1, \cdots, t_m, -t_m] t(t^2 - t_1^2) \cdots (t^2 - t_m^2)\, dt$$

for $n = 2m$ as well as for $n = 2m + 1$ function values. If the function has a $(2m + 1)$st derivative, we can write for this,

$$(23) \qquad R_n = \frac{1}{(2m + 1)!} \int_{-h}^{+h} f^{(2m+1)}(\lambda t) t(t^2 - t_1^2) \cdots (t^2 - t_m^2)\, dt,$$

where $-1 \leqq \lambda \leqq +1$.

Of course, these formulas are useful only in the case that all the roots of the equations given by the coefficient integrals are real. For $n = 8$ and $n = 10$, this is not the case. The following abscissas are given, by calculations analogous to the above example, for the function values to be chosen:

$n = 2$: $t_{\pm 1} \pm 0.5773503h$ $\qquad A = 1h,$

$n = 3$: $t_0 = 0$; $t_{\pm 1} = \pm 0.7071068h$ $\qquad A = \frac{2}{3}h,$

$n = 4$: $t_{\pm 1} = \pm 0.1875925h$; $t_{\pm 2} = \pm 0.7946545h$ $\qquad A = \frac{1}{2}h,$

$n = 5$: $t_0 = 0$; $t_{\pm 1} = \pm 0.3745414h$; $t_{\pm 2} = \pm 0.8324975h$ $\qquad A = \frac{2}{5}h,$

$n = 6$: $t_{\pm 1} = \pm 0.2666354h$; $t_{\pm 2} = \pm 0.4225187h$; $t_{\pm 3} = \pm 0.8662468h$ $\qquad A = \frac{1}{3}h.$

8. With the formulas thus far derived, it was possible to calculate exactly the integral of an integral function of nth degree, in the most favorable case, by use of n function values. There now arises the question whether or not the integral of an integral function of higher degree can

be determined exactly with the same number of values, if neither weights nor abscissas are prescribed, but if both are correspondingly determined.[7] To decide on this question, the *Lagrange interpolation formula* must first be *extended*. If we set

$$F(x) = \sum_1^n \frac{f(x_r)\varphi(x)}{(x - x_r)\varphi'(x_r)}, \tag{24}$$

then the formula 8(21) reads:

$$f(x) = F(x) + [xx_1x_2 \cdots x_n]\varphi(x). \tag{25}$$

If we now observe that $\varphi(x_1) = \varphi(x_2) = \cdots = \varphi(x_n) = 0$, it follows by use of the divided differences with repeated argument (Art. 11), and under the assumption that the given function is differentiable once, that

$$\begin{aligned} f'(x_1) &= F'(x_1) + [x_1, x_1, x_2, \cdots, x_n]\varphi'(x_1), \\ f'(x_2) &= F'(x_2) + [x_2, x_1, x_2, \cdots, x_n]\varphi'(x_2), \\ &\cdots\cdots\cdots\cdots\cdots\cdots \end{aligned} \tag{25a}$$

In general therefore,

$$[x_r, x_1, x_2, \cdots, x_n] = \frac{f'(x_r) - F'(x_r)}{\varphi'(x_r)}. \tag{26}$$

If we now apply the Lagrange interpolation formula to these n values, we obtain

$$\begin{aligned} [x_r, x_1, x_2, \cdots, x_n] = \sum_1^n \frac{((f('x_r) - F'(x_r))\varphi(x)}{(\varphi'(x_r))^2(x - x_r)} \\ + [x, x_1, x_1, x_2, x_2, \cdots, x_n, x_n]\varphi(x). \end{aligned} \tag{27}$$

By differentiation of the equation (24), it is now seen that the $F'(x_r)$ are expressed linearly by $f(x_1), f(x_2), \cdots$. If therefore the expression found here is substituted in the Lagrange formula (25), we obtain

$$\begin{aligned} f(x) &= \sum_1^n \frac{f(x_r)\cdot\varphi'(x_r)\cdot\varphi(x) + f'(x_r)(\varphi(x))^2 - F'(x_r)(\varphi(x))^2}{(\varphi'(x_r))^2(x - x_r)} \\ &\qquad + [x, x_1, x_1, \cdots, x_n, x_n](\varphi(x))^2 \\ &= \sum_1^n X_r f(x_r) + \sum_1^n \frac{f'(x_r)(\varphi(x))^2}{(\varphi'(x_r))^2(x - x_r)} \\ &\qquad + [x, x_1, x_1, \cdots, x_n, x_n](\varphi(x))^2. \end{aligned} \tag{28}$$

where the X_r are known functions of x_r, which depend neither on $f(x_r)$ nor $f'(x_r)$.

9. If this formula is integrated over the interval $-h$ to $+h$, we obtain

$$\int_{-h}^{+h} f(t)\,dt = \sum_1^n A_r f(t_r) + \sum_1^n B_r f'(t_r) \tag{29}$$
$$+ \int_{-h}^{+h} [t, t_1, t_1, \cdots, t_n, t_n](\varphi(t))^2\,dt,$$

if the variable t is again introduced. Now the abscissa values t_r can be so arranged that all the $n + 1$ weights $B_r = 0$, that therefore

$$B_r = \frac{1}{(\varphi'(t_r))^2} \int_{-h}^{+h} \frac{(\varphi(t))^2}{(t - t_r)}\,dt = 0. \tag{30}$$

This gives n equations for the unknown t_r.

If then the values

$$\frac{\varphi(t)}{(t - t_1)\varphi'(t_1)}, \frac{\varphi(t)}{(t - t_2)\varphi'(t_2)}, \cdots, \frac{\varphi(t)}{(t - t_n)\varphi'(t_n)}, \tag{30a}$$

are chosen successively for $f(t)$ in equation (29), it can be seen that the weights A_r are determined exactly by the formula used above. Then $\varphi(t)/(t - t_r)\varphi'(t_r)$ is zero for all $n - 1$ values, $t_1, t_2, \cdots, t_n$, different from t_r, while for t_r it is one. If this is substituted in the integral formula, we get

$$\int_{-h}^{+h} \frac{\varphi(t)}{(t - t_r)\varphi'(t_r)}\,dt = A_r. \tag{31}$$

The remainder term is obviously zero, since this is a case of an integral function of $(n - 1)$st degree, whose divided differences from the nth on are zero.

In addition, if we set $f(t) = \varphi(t)g_{n-1}(t)$, where $g_{n-1}(t)$ is a completely arbitrary rational integral function, then

$$\int_{-h}^{+h} \varphi(t)g_{n-1}(t)\,dt = 0, \tag{32}$$

since the $2n$ th divided difference of the function $\varphi(x)g_{n-1}(x)$ (entering into the remainder term), which is of $(2n - 1)$st degree, is zero. This equation contains the conditions $B_r = 0$, given above, as a special case. In particular, the equations (30) are satisfied if

$$\int_{-h}^{+h} \varphi(t)\, dt = 0, \int_{-h}^{+h} t\varphi(t)\, dt = 0, \tag{33}$$
$$\int_{-h}^{+h} t^2\varphi(t)\, dt = 0 \cdots \int_{-h}^{+h} t^{n-1}\varphi(t)\, dt = 0.$$

Thereupon we have *n equations, from which we can determine the n abscissa values.*

10. The left sides of these equations are nothing else than the moments from the zeroth to the $(n - 1)$st order, of the function $\varphi(t)$, and these moments can be transformed by means of the Jacobi formula [14(19)]. If we set $a = -h$, $b = +h$, $x = t$ and $\xi = 0$ in this formula, we obtain from our n equations:

$$\int_{-h}^{+h} \varphi(t)\, dt = 0,$$

$$\int_{-h}^{+h} t\varphi(t)\, dt = -h \int_{-h}^{+h} \varphi(t)\, dt + \int_{-h}^{+h} \int_{-h}^{t} \varphi(t)\, dt^2 = 0,$$

$$\int_{-h}^{+h} t^2\varphi(t)\, dt = (-h)^2 \int_{-h}^{+h} \varphi(t)\, dt + 2(-h) \int_{-h}^{+h} \int_{-h}^{t} \varphi(t)\, dt^2 \tag{34}$$
$$+ 2\cdot 1 \int_{-h}^{+h} \int_{-h}^{t} \int_{-h}^{t} \varphi(t)\, dt^3 = 0.$$

. . . .

If the value of the first is substituted in the second, the values of the first two in the third, etc., we obtain the following, in place of the above n integrals:

$$\int_{-h}^{+h} \varphi(t)\, dt = 0, \int_{-h}^{+h} \int_{-h}^{t} \varphi(t)\, dt^2 = 0, \tag{35}$$
$$\cdots \int_{-h}^{+h} \int_{-h}^{t} \cdots \int_{-h}^{t} \varphi(t)\, dt^n = 0.$$

If we now introduce the following function of $2n$th degree:

$$\Phi(t) = \int_{-h}^{t} \int_{-h}^{t} \cdots \int_{-h}^{t} \varphi(t)\, dt^n. \tag{36}$$

then this function has the property that it is zero, as well as its first n derivatives, for $t = -h$. This is self evident, since the upper and lower limits on the integral are then the same. It is also true for $t = +h$, as

follows from the n equations (35). But its nth derivative is $\varphi(t)$. The particular rational integral function of $2n$th degree which has the prescribed roots is

$$\Phi(t) = C(t^2 - h^2)^n. \tag{37}$$

Therefore, we have

$$\varphi(t) = C\frac{d^n(t^2 - h^2)^n}{dt^n}. \tag{38}$$

For $h = 1$ and $C = 1/2^n n!$ this function is known as the *Legendre polynomial or spherical function* of nth order.[8] It is customary to determine the constant so that the coefficient of the highest power is 1. Therefore we must have

$$\varphi(t) = \frac{(n)!}{(2n)!}\frac{d^n(t^2 - h^2)^n}{dt^n}. \tag{39}$$

If this function is set equal to zero and the roots of this equation are chosen as function values for the integration formula, then the condition is satisfied that all $B_r = 0$. There results the equation

$$\begin{aligned}\int_{-h}^{+h} f(t)\,dt &= \sum_1^n f(t_r)\int_{-h}^{+h}\frac{\varphi(t)}{(t - t_r)\varphi'(t_r)}\,dt \\ &\quad + \int_{-h}^{+h}[t, t_1, t_1, t_2, t_2, \cdots, t_n, t_n](\varphi(t))^2\,dt.\end{aligned} \tag{40}$$

11. It may easily be seen that *all roots of the equation* $\varphi(t) = 0$ *are real* and must lie between $-h$ and h. The equation $\Phi(t) = 0$ has $2n$ real roots, namely, an n-fold root at $+h$ and an n-fold root at $-h$. By Rolle's theorem therefore, the derivative $\Phi'(t)$ will have a real root in the interval $-h$ to $+h$, and in addition, an $(n - 1)$-fold root at the values $+h$ and $-h$. If $\Phi''(t)$ is formed and Rolle's theorem is used on both subintervals, from $-h$ or h to the root of $\Phi'(t)$, then it follows that the rational integral function of degree $2n - 2$, $\Phi''(t)$, must have two real roots in the interval $-h$ to $+h$, in addition to an $(n - 2)$-fold root at $-h$ and $+h$. If we continue step by step in this manner, we see finally that $\varphi(t) = \Phi^{(n)}(t)$ can have no roots at the limits of the integral, but that it has n real roots in the interval $-h$ to $+h$.

Let us calculate the roots and the corresponding weights, for example, in the case $n = 3$. This gives

$$\varphi(t) = \frac{3!}{6!}\frac{d^3}{dt^3}(t^2 - h^2)^3 = t^3 - \frac{3}{5}h^2 t = 0.$$

This gives the three abscissa values, $t_0 = 0$, $t_{\pm 1} = \pm h(15)^{\frac{1}{2}}/5$. From these the coefficients may be calculated by (31):

(40a)
$$A_0 = \int_{-h}^{+h} \frac{t^2 - (3/5)h^2}{-(3/5)h^2}\,dt = -\frac{5}{3h^2}\left[\frac{t^3}{3} - \frac{3}{5}h^2 t\right]_{-h}^{+h} = 2h\,\frac{4}{9},$$
$$A_1 = \int_{-h}^{+h} \frac{t^2 - (3/5)^{1/2}ht}{2\cdot(3/5)h^2}\,dt = \frac{5}{6h^2}\left[\frac{t^3}{3} - \left(\frac{3}{5}\right)^{1/2} h\frac{t^2}{2}\right]_{-h}^{+h} = 2h\cdot\frac{5}{18}.$$

The formula consequently becomes

(40b)
$$\int_{-h}^{+h} f(x)\,dx = \frac{2h}{18}\left[8f(0) + 5\left[f\left(-\frac{1}{5}(15)^{1/2}h\right) + f\left(+\frac{1}{5}(15)^{1/2}h\right)\right]\right] + R.$$

By similar calculations, the following values are found:

(40c)

$n = 1$: $t_0 = 0$,

$A_0 = 2h$,

$n = 2$: $t_{\pm 1} = \pm 0.57735027h$,

$A_{\pm 1} = h$,

$n = 3$: $t_0 = 0$, $t_{\pm 1} = \pm 0.77459667h$,

$A_0 = \frac{8}{9}h$, $A_{\pm 1} = \frac{5}{9}h$,

$n = 4$: $t_{\pm 1} = \pm 0.33998104h$, $t_{\pm 2} = \pm 0.86113631h$,

$A_{\pm 1} = 0.65214515h$, $A_{\pm 2} = 0.34785484h$,

$n = 5$: $t_0 = 0$, $t_{\pm 1} = \pm 0.53846931h$,

$t_{\pm 2} = \pm 0.90617985h$.

$A_0 = \frac{128}{225}h$, $A_{\pm 1} = 0.47862867h$,

$A_{\pm 2} = 0.23692689h$.

12. The remainder term

$$R_n = \int_{-h}^{+h} [t, t_1, t_1, t_2, t_2, \cdots, t_n, t_n](\varphi(t))^2 \, dt \tag{41}$$

can be transformed further. Since $[\varphi(t)]^2$ cannot be negative, the mean value theorem of integral calculus can be applied to this integral. If we designate by τ an intermediate value of the $2n$th difference in the interval $-h$ to $+h$, we have

$$R_n = [\tau, t_1, t_1, t_2, t_2, \cdots, t_n, t_n] \int_{-h}^{+h} (\varphi(t))^2 \, dt. \tag{42}$$

The integral

$$J = \int_{-h}^{+h} (\varphi(t))^2 \, dt = \frac{((n)!)^2}{((2n)!)^2} \int_{-h}^{+h} \frac{d^n(t^2 - h^2)^n}{dt^n} \cdot \frac{d^n(t^2 - h^2)^n}{dt^n} \, dt \tag{43}$$

can be transformed by repeated integration by parts (cf. 27.9), so that we finally obtain as the remainder term

$$R_n = \frac{(2h)^{2n+1}}{2n + 1} \frac{((n)!)^4}{((2n)!)^2} [\tau, t_1, t_1, t_2, t_2, \cdots, t_n, t_n]. \tag{44}$$

If the function to be integrated has a $2n$th derivative, and if this is continuous, then we can express the $2n$th divided difference by it and obtain the usual form of the remainder term

$$R_n = \frac{(2h)^{2n+1}}{2n + 1} \frac{((n)!)^4}{((2n)!)^3} \cdot f^{2n}(\bar{\tau}), \tag{45}$$

where $-h \leqq \bar{\tau} \leqq h$. In the formulas given above therefore, we have the following remainder terms:

$$\begin{gathered} R_1 = \frac{2h^3}{3} f''(\bar{\tau}), \quad R_2 = \frac{h^5}{135} f^{(4)}(\bar{\tau}), \quad R_3 = \frac{h^7}{15750} f^{(6)}(\bar{\tau}), \\ R_4 = \frac{h^9}{3472875} f^{(8)}(\bar{\tau}), \quad R_5 = \frac{h^{11}}{1237732650} f^{(10)}(\bar{\tau}). \end{gathered} \tag{45a}$$

13. Bortkiewitz[9] has investigated the usefulness of formulas with equidistant ordinates, particularly those which do not approximate the interval with a single curve, but which use several curves over various subintervals. He worked with formulas such as we have used in the preceding articles. This included the *integration of empirical functions*, such as the determination of the mean life expectancy, from mortality tables, for the population of Berlin in the years 1800-1900. Bortkiewitz shows that formulas with negative weights (as occur in the MacLaurin formula for

$n = 7$) are to be avoided, and that the accuracy criteria as they were developed here sometimes fail. The reason for this is that the chief assumption—that the divided differences of a particular order become small throughout the interval, i.e., that the function may be approximated by a rational integral function of corresponding order—is not necessarily valid.

14. The Gauss method, as well as that of Tschebyscheff, assumes that the function values can be determined exactly for the calculated abscissa values. Since in general computations are performed with finite decimal fractions, this is usually not the case. The indicated accuracy of the Gauss formula is therefore illusory. If the function to be integrated is developed in a series, then, in the formation of the approximation value in the correction term, the factors of the nth to $(2n - 1)$st derivatives become very small, but do not vanish. More accurate investigations have been given in this case by Moors.[10]

The formulas specified here are not the only ones which have been developed. There are formulas which prescribe a portion of the abscissa for which the function values are to be determined, and so determine the rest, including the collected weights, that the correction is made a minimum, i.e., with differentiable functions, the correction term contains the mean value of a derivative of as high an order as possible.[11] For example, Lobatto[12] has developed formulas for the case in which the end ordinates of the integration interval are used.[13] Development of all these formulas would require too much space here.

Finally, we observe that the mean value methods can also be extended to functions of several variables. In such a case it is natural that the position of the function value to be chosen depends on the boundaries of the region.[14]

NOTES

1. Newton, *Philosophiae naturalis principia mathematica* (1687) 3, Prop. XL, Lemma 5.
2. Cotes, *Harmonia mensurarum* (1722). (Appendix).
3. Moors, *Valeur approximative d'une intégrale définie* (Paris, 1905), gives the values up to n = 13; also the numerical values are given there for the formulas to be mentioned below.
4. Steffensen, *Skandinavisk Aktuariestidskrift* (1921), p. 201; *Interpolationslaere* (Copenhagen, 1925), Art. 16.
5. Walter, *Skandinavisk Aktuariestidskrift* (1925), p. 148.
6. Tschebyscheff, *Journal de math.* (2) 19, (1874), pp. 19-34.
7. Gauss, *Com. soc. reg. scient. Gottengensis* III (1826), Werke II, pp. 163-196; Encke, *Berliner Astronomisches Jahrbuch* (1863). *Gesammelte Abhandlungen* I (Berlin, 1888), pp. 100-124.
8. Legendre, *Mémoires de Math. et Phys.* (10) (Paris, 1785); Heine, *Theorie der Kugelfunktionen*, 2nd ed. (Berlin, 1878).

9. Bortkiewitz, Skandinavisk Aktuarietidskrift (1926), p. 1.
10. Moors, *Valeur approximative d'une intégrale définie* (Paris, 1905).
11. Christoffel, *Journal f. Math.* 55 (1858), p. 66.
12. Lobatto, *Lessen over Integral-Rekening*, (La Haye, 1852), Art. 207.
13. Radau, *Journal de math.* (3) 6 (1880), p. 283.
14. *Math. Enzyklopädie* II C 2, No. 13b and c.

17. The Planimeter.[1]

1. Devices which serve to determine the value of a definite integral mechanically by a tracing of the curve which represents the function to be integrated, are known as planimeters. They consist essentially of a rod, the tracing arm, on one end of which is mounted the tracing pen which follows the curve. The other end of this rod is moved by some mechanism on a straight line or on a circular path, by means of a second rod, the pole arm, which can be turned about an endpoint, the pole, and which is joined to the other end of the tracing arm by means of a hinge. In the first case the apparatus is called a linear planimeter, in the second, a polar planimeter.

2. To develop the theory of the planimeter, we compute (according to Rothe[2]) the area which is enclosed by the tracing arm. If we denote the length of the tracing arm by l, the coordinates of the endpoint F, which is the tracing point, by x, y, the coordinates of the other endpoint Q by ξ, η, and the angle between the tracing arm and the x axis by θ, then, by Fig. 53,

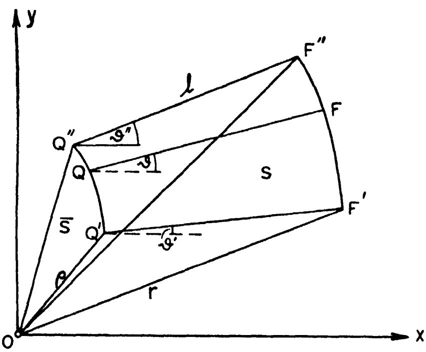


FIG. 53

$$x - \xi = l \cos \theta, \qquad y - \eta = l \sin \theta. \tag{1}$$

In this case, for a given motion, we can regard x, y, ξ, η, as well as l all as function of one variable t. This variable may well be the time, because both endpoints Q and F of the rod are moved rigidly on the two curves $Q'Q''$, $F'F''$, so that the entire motion has only one degree of freedom. If the initial position of the tracing arm $Q'F'$ is determined by the parametric value t', the final position $Q''F''$ by t'', then sectors S and $\bar{S}$ are

swept out by the radius vectors r and ρ. The values r and ρ correspond to the distances from the two endpoints of the tracing arm to the origin of an arbitrarily chosen coordinate system. These areas are measured by the integrals

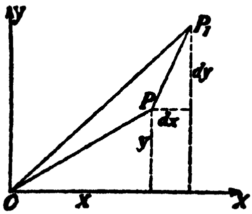


FIG. 54

$$(2)\qquad 2S = \int_{t'}^{t''} (x\,dy - y\,dx),\qquad 2\bar{S} = \int_{t'}^{t''} (\xi\,d\eta - \eta\,d\xi).$$

The differential for the area of the sector can easily be read from Fig. 54:

$$(3)\qquad \begin{aligned} 2dS &= (x + dx)(y + dy) - xy - 2y\,dx - dy\cdot dx \\ &= (x\,dy - y\,dx). \end{aligned}$$

From the above equations, we now introduce

$$(4)\qquad \begin{aligned} x &= \xi + l\cos\theta, & y &= \eta + l\sin\theta, \\ dx &= d\xi - l\sin\theta\,d\theta, & dy &= d\eta + l\cos\theta\,d\theta, \end{aligned}$$

from which we obtain, by multiplication,

$$(5)\qquad \begin{aligned} x\,dy - y\,dx = \xi\,d\eta - \eta\,d\xi &+ l^2(\cos^2\theta + \sin^2\theta)\,d\theta \\ &+ l(\xi\cos\theta + \eta\sin\theta)\,d\theta + l(\cos\theta\,d\eta - \sin\theta\,d\xi), \end{aligned}$$

so that

$$(6)\qquad \begin{aligned} 2S = 2\bar{S} + l^2 \int_{t'}^{t''} d\theta &+ l\int_{t'}^{t''} (\xi\cos\theta + \eta\sin\theta)\,d\theta \\ &+ l\int_{t'}^{t''} (\cos\theta\,d\eta - \sin\theta\,d\xi). \end{aligned}$$

If the angles of the initial and final positions of the tracing arm with respect to the x axis are θ' and θ'', the first integral becomes

$$(7)\qquad l^2 \int_{t'}^{t''} d\theta = l\,(\theta'' - \theta').$$

To transform the latter integral, integration by parts is applied to the parts of the preceding integral. We then get

$$(8)\qquad \begin{aligned} \int_{t'}^{t''} (\xi\cos\theta + \eta\sin\theta)\,d\theta &= [\xi\sin\theta - \eta\cos\theta]_{t'}^{t''} \\ &+ \int_{t'}^{t''} (\cos\theta\,d\eta - \sin\theta\,d\xi). \end{aligned}$$

The first member of the right side must be transformed still further. If the angle which the radius vector ρ, from 0 to Q, forms with the positive direction of the x axis is denoted by φ, then $\xi = \rho \cos \varphi$, $\eta = \rho \sin \varphi$. Therefore we get

$$
\begin{aligned}
(9) \qquad [\xi \sin \theta - \eta \cos \theta]_{t'}^{t''} &= \rho[\cos \varphi \sin \theta - \sin \varphi \cos \theta]_{t'}^{t''} \\
&= [\rho \sin (\theta - \varphi)]_{t'}^{t''}.
\end{aligned}
$$

If the distance from the tracing arm to the origin is denoted by p, then we can infer from Fig. 55 that

$$(10) \qquad [\xi \sin \theta - \eta \cos \theta]_{t'}^{t''} = [p]_{t'}^{t''} = p'' - p',$$

where the primes again indicate the initial and final positions of the tracing arm.

If all this is now substituted in the equation (6) for S, we have

$$
\begin{aligned}
S &= \bar{S} + \frac{l^2}{2}(\theta'' - \theta') \\
(11) \qquad &\quad + \frac{l}{2}(p'' - p') + l \int_{t'}^{t''} (\cos \theta \, d\eta - \sin \theta \, d\xi).
\end{aligned}
$$

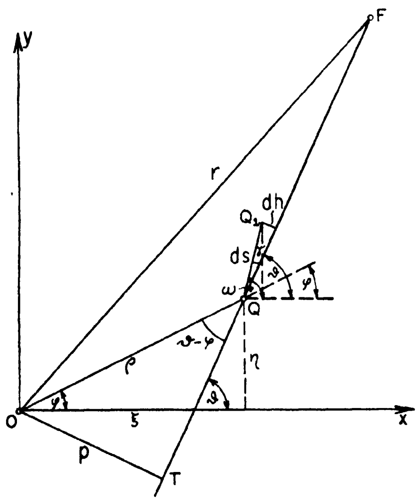


FIG. 55

The final integral is also to be transformed further. From Fig. 55, in which QQ_1 represents the path differential of the one endpoint Q of the tracing arm, it may be deduced that

$$d\xi = ds \cos \omega, \qquad d\eta = ds \sin \omega.$$

By use of the relations introduced in Fig. 55, the last integral becomes

$$(12) \qquad \begin{aligned} \int (\cos \theta \, d\eta - \sin \theta \, d\xi) &= \int (\cos \theta \sin \omega - \sin \theta \cos \omega) \, ds \\ &= \int \sin (\omega - \theta) \, ds = \int \sin \gamma \, ds = \int dh. \end{aligned}$$

Now the $\int dh$ measures the displacement of the end point Q of the tracing arm perpendicular to its direction. If this value is also substituted in the above equation, we obtain the *general planimeter equation*:

$$(13) \qquad S = \overline{S} + \frac{1}{2} l^2(\theta'' - \theta') + \frac{1}{2} l(p'' - p') + l \int_{t'}^{t''} dh.$$

3. Devices are mounted on the planimeter which measure the integral $\int dh$, i.e., the displacement of the endpoint Q perpendicular to the direction of the tracing arm. The mechanism most frequently used for this purpose is the *integrating wheel.* This is a wheel with milled edges, which moves by sliding in the direction of its axis, parallel to the tracing arm. In the case of displacement perpendicular to the tracing arm, and therefore perpendicular to its axis on the drawing plane, the wheel rolls. In all other displacements, it partly slides and partly rolls about a displacement of the tracing arm perpendicular to its own direction. From Fig. 56 we see that in the displacement of the contact point of the wheel a distance Δs in a direction which makes the angle γ with the wheel axis, the wheel itself rolls a distance

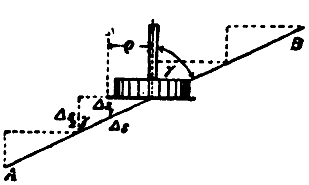


FIG. 56

$$\Delta s_1 = \Delta s \sin \gamma = \Delta u 2\rho\pi.$$

Here Δu is the fraction of one revolution through which the integrating wheel, with radius ρ, turns during the course of the motion.

The number of complete revolutions of the integrating wheel is read off on a dial moved by an endless screw. The subdivisions of these revolutions are read off the circumference of the wheel itself by means of a vernier, divided into 100 parts. Readings may then be made to 1/1000 of a revolution.

If AB is a curve and γ is the angle the wheel axis makes with the tangent to the curve, then the number of revolutions turned in the tracing of the contact point of the integrating wheel along this curve is

$$u = \frac{1}{2\rho\pi}\int \sin\gamma\, ds = \frac{1}{2\rho\pi}\int dh, \tag{14}$$

if the wheel is placed at the end point Q of the tracing arm.

For technical reasons, this is usually not the case. More frequently the integrating wheel is at some position adjacent to the tracing arm, so that the axis parallel to the tracing arm is at a distance d. Also, the plane of the wheel perpendicular to the drawing plane cuts the tracing arm at a distance c from the endpoint, so that the projection of the radius vector from Q to the contact point of the tracing arm forms the angle δ.

Each motion of the tracing arm may then be considered as composed of three partial motions (Fig. 57):

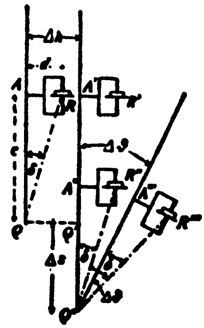


FIG. 57

(a) Displacement of QA to $Q'A'$, a distance Δh perpendicular to the tracing arm; rotation of the wheel through $\Delta u_1 = \Delta h/2\pi\rho$.

(b) Displacement of $Q'A'$ to $Q''A''$ in the direction of the tracing arm; no rotation of the wheel;

(c) Rotation of the tracing arm about Q from $Q''A''$ to $Q''A'''$. Movement of the contact point of the wheel on the circle $R''R'''$ of radius $c/\cos\delta$ and of length $c\Delta\theta/\cos\delta$. Since the wheel axis makes an angle $90 + \delta$ with the direction of motion, the wheel turns through

$$\Delta u_2 = \frac{c\Delta\theta}{\cos\delta}\cdot\frac{\sin(90+\delta)}{2\rho\pi} = \frac{c}{2\rho\pi}\cdot\Delta\theta. \tag{15}$$

The rotation of the wheel is therefore independent of the distance of the wheel axis from the tracing arm. The total rolling of the wheel is then

$$u = \int_{t'}^{t''} du_1 + du_2 = \frac{1}{2\rho\pi}\left(\int_{t'}^{t''} dh + c(\theta'' - \theta')\right). \tag{16}$$

If this value is substituted for $\int dh$ in the general planimeter equation (13), we get the *equation for a planimeter with the integrating wheel on static support*:

$$S = \overline{S} + \left(\frac{l^2}{2} - lc\right)(\theta'' - \theta') + \frac{l}{2}(p'' - p') + 2l\rho\pi u. \tag{17}$$

4. The planimeter most widely used is the *polar planimeter* developed by Amsler[3], in which one end Q of the tracing arm A is moved on a circle by means of the polar arm P. The details of the planimeter are shown in Fig. 58 which reproduces the so-called compensation planimeter. In this

the connection between the tracing arm and the polar arm is made by a ball joint, which can be dismantled. In this way the error of the hinge inclination, i.e., the error which results from the fact that the joint is not perpendicular to the plane of the drawing, is avoided. Besides, the

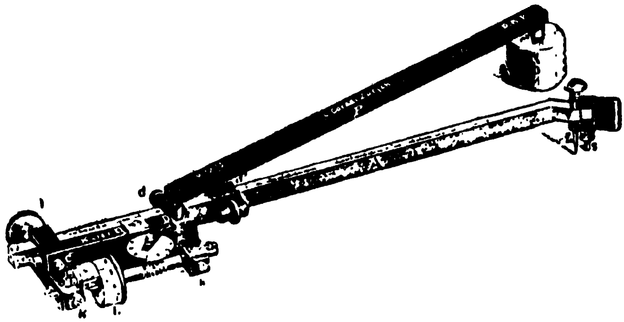


FIG. 58

planimeter can be dismantled at the junction of the rod P and the tracing pen f, in the ball joint. We can then move P through to the other side, and so carry out the tracing of the area to be measured with two different positions of the planimeter. If we form the mean of the two measurements which are then obtained, we reduce a second error of the instrument, namely, that the axis of the integrating wheel is not parallel to the tracing arm.[4]

To measure small surfaces with a planimeter, the pole is so set that it

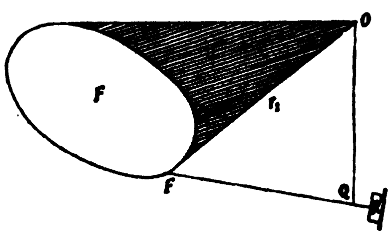


FIG. 59

is outside of the surface to be measured. The shaded area in Fig. 59 is traced over by the radius vector in one direction and then in the other, in girdling the area F with the tracing point. The area is then canceled out. Furthermore, in a complete circuit of the area, the planimeter is brought back to its initial position. Consequently $\theta'' = \theta'$ and $p'' = p'$ and the radius vector OQ, here the pole arm, sweeps over the same sector in a positive and negative sense, so that $\overline{S} = 0$. The planimeter equation is then

$$F = 2\rho\pi \cdot l \cdot u. \tag{18}$$

The position of the measuring wheel is then immaterial, provided that its axis is parallel to the tracing arm.

The factor $2\pi\rho l$ with which the number of revolutions is to be multiplied, is known as the *planimeter constant.* The simplest way of determining this constant is the following. A moderately sized area of known magnitude—say a square 10 cm. on a side—is traversed, and the area of this surface is divided by the number of turns of the wheel. The resultant value is the planimeter constant. We form the mean value of this quantity from several circuits in the direction in which we intend to perform the circuit in the actual measurement. The constant is so chosen for most polar planimeters that one complete revolution of the integrating wheel corresponds to about 1 square meter. Before the final measurement, the surface whose area is to be measured is traversed in an approximate

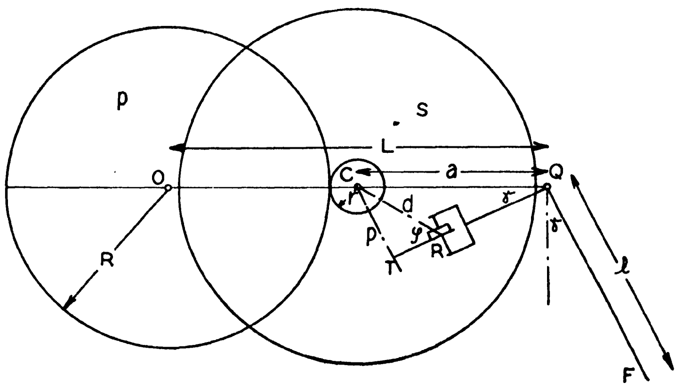


FIG. 60

fashion, in order to ascertain whether the wheel stays on the paper. In the circuit, the tracing point is moved, as far as possible, so that it is viewed in the direction of motion, in order to avoid any possible parallax. The number of revolutions is equal to the difference of the readings before and after the circuit. Tracing with the aid of a rule along a straight portion of the curve is to be avoided, since the errors in free tracing, resulting from back and forth oscillations, cancel each other.

For large areas, the pole is placed within the area. Then for a complete circuit, $P'' = P'$ again, but here, $\theta'' - \theta' = 2\pi$, $\bar{S} = L^2\pi$, where L is the length of the polar arm. In this case the planimeter equation becomes

$$F = (L^2 + l^2 - 2lc)\pi + 2\pi\rho \cdot lu. \tag{19}$$

Greater accuracy is achieved with planimeters in which the support,

on which the integrating wheel rolls, is also moved. This may be accomlished by a rolling on a polar disk, as in the *disk planimeter.*[5] This planimeter uses the so-called *Gonella integrating mechanism* (1824). A wheel C rolls on the polar disk P of radius R. This polar disk has milled edges. The axis of C, perpendicular to the drawing plane, is moved by the polar arm OQ, and is rigidly connected to the disk S. The integrating wheel R rests on this disk at a distance d from the axis, and is borne by an arm QR perpendicular to the tracing arm l, so that its plane passes exactly through the endpoint Q of the tracing arm. Therefore $c = 0$.

If now the polar arm is rotated through $\Delta\chi$, the disk, as a consequence of the rolling of C on the polar plane, rotates through $\Delta\psi = (R/r)\Delta\chi$, and the contact point of the integrating wheel is moved through $d(\Delta\psi) = (Rd/r)\Delta\chi$. If the angle between the plane of the wheel and d is denoted by φ, then the measuring wheel rotates through

$$(20) \qquad 2\rho\pi\Delta u = \frac{Rd}{r}\sin\varphi\Delta\chi = \frac{Ra}{r}\sin\gamma\cdot\Delta\chi = \frac{R}{r}(L - (r + R))\sin\gamma\cdot\Delta\chi,$$

as may be deduced from Fig. 60. Now $L(\Delta\chi)$ is the path traversed by the point Q in this motion, and $(L\Delta\chi)\sin\gamma$ is the component of this

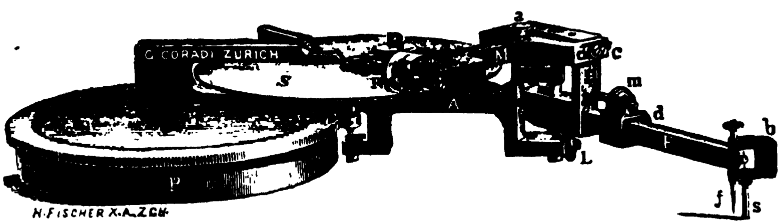


FIG. 61

motion perpendicular to the tracing arm, which we have denoted by Δh. Therefore we have

$$(21) \qquad 2\rho\pi\Delta u = \frac{R}{r}\left(1 - \frac{r + R}{L}\right)L\Delta\chi\sin\gamma = \frac{R}{r}\left(1 - \frac{r + R}{L}\right)\Delta h,$$

which, upon integration, is

$$(22) \qquad 2\rho\pi u = \frac{R}{r}\left(1 - \frac{r + R}{L}\right)\int_{t'}^{t''}\Delta h.$$

If we introduce this in the general planimeter equation and observe that

for the choice of the pole outside of the surface to be measured, $P'' = P'$, $\theta'' = \theta'$, $\overline{S} = 0$, and

$$F = \frac{2rL \cdot l \cdot 2\rho\pi}{R(L - (r + R))} u \tag{23}$$

for a complete circuit of the closed surface. If the pole were placed inside, we would still get $(L^2 + l^2)\pi$. Here also the planimeter constant is determined by a circuit of a surface of known area. The dimensions of the instrument are so chosen that one complete revolution of the wheel corresponds to about 5-20 cm.^{2}, according to the length l of the tracing arm, which is adjustable.

Such planimeters have also been constructed as linear planimeters. The whole instrument is then moved on two wheels in the direction of the x axis. Then Q moves on a parallel to the x axis, and the rotation of the disk is proportional to the displacement in the direction of the x axis.

6. Another integrating device is the *spherical wheel planimeter*, which is usually constructed as a linear planimeter. Here the disk is replaced by a spherical cap K of radius ρ_1 and the integrating wheel is replaced by a cylinder of radius ρ_2, which is pressed against the spherical cap by means of a spring. The whole apparatus travels on rollers of radius R in the

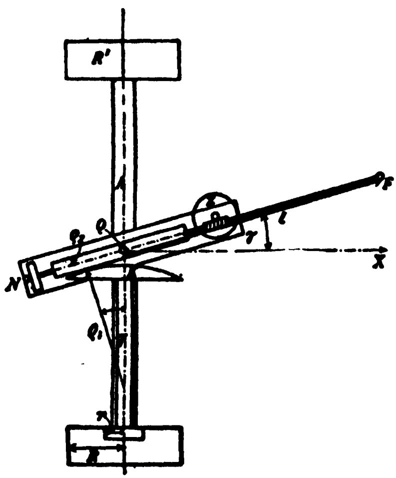


FIG. 62

direction of the x axis by means of these rollers and a small milled wheel of radius r. In a movement of the apparatus a distance Δr, the spherical cap is turned through an angle $\Delta x/r = \Delta\varphi$. If the notation of Fig. 62 is used, the rotation u of the cylinder is given by

$$(24) \qquad 2\rho_2\pi\Delta u = \rho_1 \sin\gamma\cdot\Delta\varphi = \frac{\rho_1 \sin\gamma\Delta x}{r} = \frac{\rho_1\Delta h}{r},$$

since $\Delta x \sin \gamma$ is the displacement of the tracing arm FQ perpendicular to its axis. Therefore

$$(25) \qquad u = \frac{\rho_1}{2\rho_2\pi\cdot r}\cdot\int_{t'}^{t''} dh,$$

and if this is substituted in the planimeter equation, we get, for a complete circuit of the surface,

$$(26) \qquad F = \frac{2\rho_2\pi r l}{\rho_1} u.$$

The planimeter constant of the spherical wheel planimeter is just about as large as that of the disk planimeter. Each linear planimeter has the advantage that surfaces of arbitrary length can be measured; naturally, the breadth of the curve is limited by the length of the tracing arm. An especially simple planimeter is the type known as the *Prytz* or *hatchet planimeter.*[6]

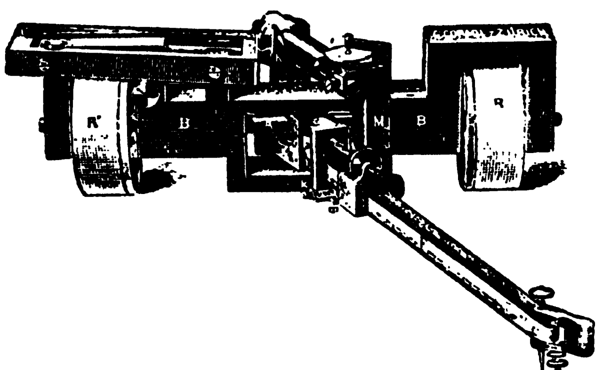


FIG. 63

7. We shall now give a few *examples of the application of planimeters.*

a. In *orometry* the problem is to find the mean height of a certain bounded region. That is, we seek the volumes of irregularly bounded bodies. The shapes of these bodies are given by contour lines. If z is the height above the zero level, and $g(z)$ is the area of the surface bounded by the contour line at the height z, then the mean height is

$$h = \frac{1}{g(0)}\int_0^z g(z)\, dz.$$

The individual values of the function $g(x)$ are then to be found by use of the planimeter. If the points so obtained are plotted as a function of z, the end points are connected by a smooth curve, and the area is taken with a planimeter, the value of the integral is obtained.

Another example of this type is the following: a scale is to be prepared for an artificial lake, for which shore contours are given, from which the volume of water contained behind the dam can be read off as a function of the water level.[7]

b. The *static moment*[8] and the *center of mass* of a surface can be found by graphing the data and taking the area of the new surface. The static moment about the y axis is

$$S_x = \int yx\,dx = \frac{1}{2}\int yd(x^2) = \frac{1}{2}\int y\,d\xi, \tag{27}$$

where $x^2 = \xi$. The curve is drawn (using a quadratic scale $\xi = x^2$ on the x axis) so that the ordinate belonging to x is changed to ξ. This can be done without construction of the scale. We can determine $\xi = x^2$ with

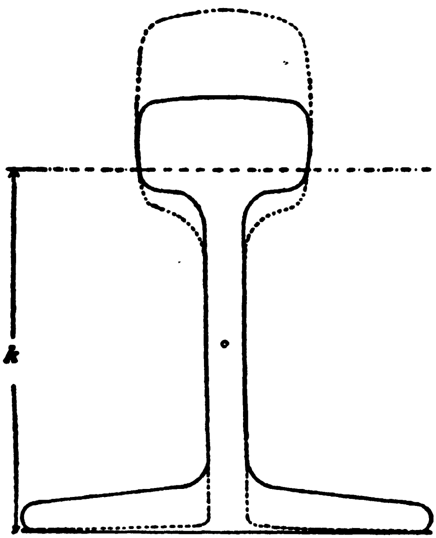


FIG. 64

a slide rule and enter the value of y at ξ. Then the graphing can be carried out in a comparatively short time. By use of the planimeter on the resultant surface, we obtain $2S_x$. In the same way, we obtain

$$S_y = \int xy\,dy = \frac{1}{2}\int xd(y^2) = \frac{1}{2}\int x\,d\eta.$$

For the symmetrical rail cross section in the accompanying Fig. 64, we get $F = 7.775$ cm.^{2} for the surface area. The drawing is not on a scale $\eta = y^2$, but follows the scale $\eta = y^2/5$. We get 8.2 cm.^{2} for the surface area of the dotted figure, so that the static moment about the lower edge of the cross section is $S_y = 5 \times 8.2/2$ cm.^{3}

Now the coordinates of the center of mass are

$$(28) \qquad x_s = \frac{S_x}{F} = \frac{\int xy\,dx}{\int y\,dx}; \qquad y_s = \frac{S_y}{F} = \frac{\int yx\,dy}{\int x\,dy}.$$

In the above figure, the center of mass naturally lies on the axis of symmetry, and is

$$y_s = \frac{20.5 \text{ cm}^3}{7.775 \text{ cm}^2} = 2.65 \text{ cm}$$

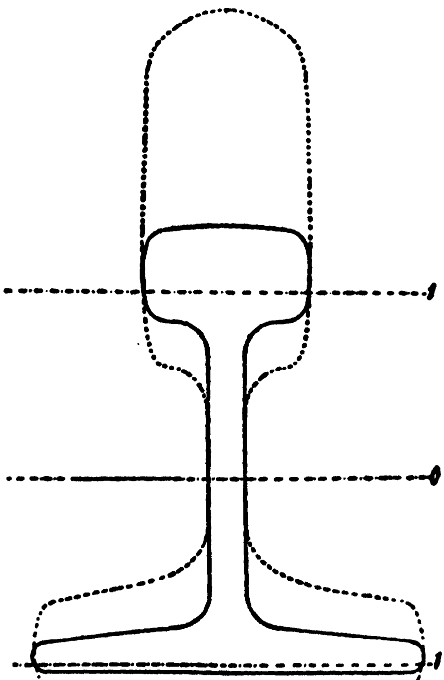


FIG. 65

distant from the lower edge.

c. The *axial moment of inertia* about the y axis is

$$T_x = \int yx^2\,dx = \frac{1}{3}\int y\,dx^3 = \frac{1}{3}\int y\,d\xi. \tag{29}$$

We must therefore draw the curve to the scale $\xi = x^3/3$, and measure the area with a planimeter. Just as with the static moment, a change in scale is usually necessary.

For example, the scale $\eta = y^3/3.125$ is chosen in Fig. 65, in which the moment of inertia about an axis parallel to the x axis, through the center of mass, is determined. The value obtained with a planimeter is then to be multiplied by 3.125. We get

$$T_y = \int xy^2\,dy = \frac{1}{3}\int x\,d\eta = \frac{3.125}{3} \times 18.98 = 19.75\text{ cm}^4.$$

d. The *product of inertia* of a surface is

$$T_{x,y} = \iint xy\,dx\,dy = \frac{1}{4}\int y^2\,dx^2 = \frac{1}{4}\int \eta\,d\xi. \tag{30}$$

Here both the coordinates must be transformed to new scales.

e. The *moment of inertia about an axis* perpendicular to the surface is

$$T_p = \iint r^2\,do = \iint r^3\,dr\,d\varphi = \frac{1}{4}\int r^4\,d\varphi = \frac{1}{4}\int R^2\,d\varphi. \tag{31}$$

If we write $r^2 = R$, we can regard the latter integral as a surface and, after drawing, determine T_p with the planimeter.

f. Let us give the corresponding formulas for solids of rotation. If x is the axis of rotation, the volume is

$$V = \pi\int_{x_1}^{x_2} y^2\,dx = \pi\int_{x_1}^{x_2} \eta\,dx. \tag{32}$$

The necessary drawing of the meridian curve is $\eta = y^2$ (cf. 14.9).

The static moment with respect to the yz plane is

$$\begin{aligned} S_x &= \iiint x\,dv = \int_0^{2\pi}\int_{x_1}^{x_2}\int_0^{y} x\cdot y\,d\varphi\,dx\,dy \\ &= 2\pi\int_{x_1}^{x_2} \frac{y^2}{2}\frac{d(x^2)}{2} = \frac{\pi}{2}\int_{\xi_1}^{\xi_2} \eta\,d\xi. \end{aligned} \tag{33}$$

A drawing on two new scales $\eta = y^2$, $\xi = x^2$ is necessary.

The moment of inertia about the axis of rotation

$$(34)\qquad \begin{aligned} T_x = \int y^2\,dV &= \int_0^{2\pi}\int_{x_1}^{x_2}\int_0^{y} y^2 y\,d\varphi\,dx\,dy \\ &= 2\pi\int_{x_1}^{x_2}\frac{y^4}{4}\,dx = \frac{\pi}{2}\int_{x_1}^{x_2}\eta\,dx. \end{aligned}$$

The ordinate is to be drawn $\eta = y^4$ with an unchanged abscissa. Then the above quantities are determined with the planimeter.

g. Rothe gives further examples, e.g., the determination of the mean light intensity of a light source[9] and the determination of the probability of hitting a target.[10].

The determination of the static moment and the axial moment of inertia of a surface can also be carried out without redrawing, by means of a moment planimeter.[11]

NOTES

1. Cf. J. Groeneveld, *Z. f. Instrumentenkunds* 47 (1927), p. 1, 113, 185.
2. From his lectures. Schrutka, *Österr. Z. f. Vermessungsw.* (1927), No. 5.
3. Amsler, *Vierteljahrsschrift der naturforschenden Gesellschaft* (Zurich, 1856), 1.
4. Willers, *Mathematische Instrumente* (Berlin, 1926), pp. 72-73.
5. Coradi, *Z. f. Vermessungswesen,* 10 (1881).
6. Prytz, *Tekniske Forenings Tidsskrift* (1886); *Tidsskrift for Opmaalings- og Matrikulavesen* (1895). Willers, *op. cit.*, Art. 13.
7. Rothe, *Elektrotechnische Zeitschrift,* 41 (1920), pp. 999-1002.
8. Nehls, *Zivilingenieur* (1874); cf. 14.10 also.
9. Rothe, *loc. cit.*
10. Rothe, *Artilleristische Monatschefte* (1916), 110, pp. 65-91; 111, pp. 125-154.
11. Willers, *op. cit.*, Art. 14.

CHAPTER FOUR

PRACTICAL EQUATION THEORY

18. Approximate Calculation of the Roots of any Equation.

1. There are three general methods for the determination of the roots of any function with one variable, or of several functions with the corresponding number of variables: the method of false position, the method of iteration, and the Newton method of approximation.

With the *method of false position*, linear interpolation is used for the determination of the roots. If we draw the graph of the function $y = f(x)$, the roots of which we are seeking, the intersection of this curve with the x axis gives the desired value of the root. If we now have two values x_1 and $\bar{x}_1$, for which the function has opposite signs, and if the function is continuous between these values, then there must be at least one root, x_0, between them. If $f'(x)$ does not change sign in the interval $x_1 < \bar{x}_1$, then there must be only one root there. If $f'(x)$ does not vary too rapidly in the interval, then the curve can be approximated by a straight line through the points x_1, y_1 and $\bar{x}_1$, $\bar{y}_1$, in exactly the same way as was used with linear interpolation in tables. This straight line has the equation

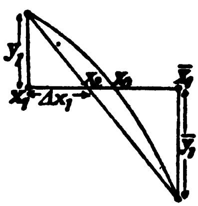


FIG. 66

$$y - y_1 = \frac{\bar{y}_1 - y_1}{\bar{x}_1 - x_1}(x - x_1).$$

The x intercept, which gives a better approximation of the desired root than was obtained with x_1 or $\bar{x}_1$, is obtained for $y = 0$. The correction $\Delta x_1 = x - x_1$ to be added to x_1 is then

$$\Delta x_1 = -\frac{\bar{x}_1 - x_1}{\bar{y}_1 - y_1}\, y_1 . \tag{1}$$

If $f''(x)$ does not change its sign in the observed interval, then the new approximation value lies on the concave side of the curve.[1] If for the value x_2 so obtained, $f(x_2)$ has the same sign as $f(x_1)$, then a larger value $\bar{x}_2$ is considered, which lies within the old interval, so that $f(\bar{x}_2)$ has the opposite sign to $f(x_2)$. If $f(x_1)$ and $f(x_2)$ have opposite signs, we choose $\bar{x}_2 < x_2$ so that $f(x_2)$ and $f(\bar{x}_2)$ have opposite signs. We then proceed as above and obtain the new correction

$$\Delta x_2 = -\frac{\bar{x}_2 - x_2}{\bar{y}_2 - y_2} y_2 .$$

This process is continued. The narrower the interval and the smaller the variation of $f'(x)$, the better the curve is approximated by a straight line, and so much more rapidly does the process converge. In general, the method of false position necessitates a great deal of calculating, and is only practical if the functions concerned are tabulated.[2]

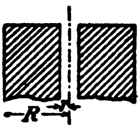


FIG. 67

2. *Example* (from Rothe). For high potential conduction through walls, tubular insulators are used which are covered with metal on the inner and outer surfaces. The ratio q of the line voltage to the maximum admissible field strength (breakdown potential) is given by $q = r \ln R/r$ (Deutscher Kalender f. Elektrotechniker 1917, I, p. 90). What ratio must the external diameter $2R$ have to the bore width in order that the cross section Q be a minimum?

If we set $R/r = x$, then $r = q/\ln x$; therefore

$$y = Q/\pi = (R^2 - r^2) = r^2(x^2 - 1) = q^2(x^2 - 1)/(\ln x)^2$$

must be a minimum. The first derivative must then be zero:

$$y^1/q^2 = \frac{d}{dx}\left(\frac{x^2 - 1}{(\ln x)^2}\right) = \frac{2x}{(\ln x)^2} - \frac{2(x^2 - 1)}{x(\ln x)^3} = 0.$$

Neither x nor $\ln x$ is zero or infinite; therefore,

$$\ln x - \left(1 - \frac{1}{x^2}\right) = 0 \qquad \text{or} \qquad x - e^{1-(1/x^2)} = 0.$$

Since e^x is tabulated, e.g., by Hayashi,[3] we use the second form and calculate a few values first, in order to get a general picture.

TABLE I.

x	1	1.5	2	2.5
$x - e^{1-(1/x^2)}$	0	−0.242	−0.117	0.185

There must then be a root between 2 and 2.5.

$$\Delta x_1 = \frac{+0.5}{0.302} \times 0.117 = 0.194 \approx 0.2.$$

If we calculate further, we must take, for $x_2 = 2.2$, and since $f(x_2)$ has the same sign as $f(x_1)$, $\bar{x}_2 > x_2$. We have

TABLE II.

x	$\frac{1}{x^2}$	$1 - \frac{1}{x^2}$	$e^{1-(1/x^2)}$	y
2.2	0.2065	0.7935	2.211123	−0.011123
2.25	0.1972	0.8028	2.231782	+0.018218

From this we obtain

$$\Delta x_2 = \frac{0.05}{0.02933} \times 0.0111 = 0.0189.$$

Therefore $x \approx 2.219$.

3. If we want *to transform the method of false position to two equations with two variables*[4], we must choose the representation of the function in scale form instead of a curve. The scale of y corresponds then to the scale of x. To the interval $\bar{x}_n - x_n$ there corresponds on this scale an interval $\bar{y}_n - y_n$ in which lies the origin of the scale. This divides the length in the same ratio in which one must divide the length $\bar{x}_n - x_n$ in order to get a better approximation for the root. If we have to find the roots of two equations $\varphi(x, y) = 0$; $\psi(x, y) = 0$ with two unknowns, then we set

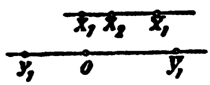


FIG. 68

$$X = \varphi(x, y), \qquad Y = \psi(x, y), \tag{2}$$

and map the x, y plane on an X, Y plane. If x_0, y_0 are the values for which the functions are zero, and if the two functions are regular in the neighborhood of these values, we can develop X, Y in powers of $(x - x_0)$, $(y - y_0)$. In these developments the constant terms vanish, but the linear coefficients are different from zero. Therefore we get an approximate mapping of a small region in the vicinity of x_0, y_0, if we break off the development with the linear terms, i.e., if we set:

$$X = a_1(x - x_0) + b_1(y - y_0); \qquad Y = a_2(x - x_0) + b_2(y - y_0). \tag{3}$$

But this defines an affine mapping. Therefore three points p_1, p_2, p_3 of the xy plane correspond to three points P_1, P_2, P_3 of the XY plane. The desired point p_0 now lies so near to the three points p_1, p_2, p_3 that the quadrangle $p_0p_1p_2p_3$ is affine to the quadrangle $OP_1P_2P_3$. Since with an affine mapping, the length ratios remain invariant, we draw the two lines P_1Q_1 and P_2Q_2 in the triangle $P_1P_2P_3$ and divide the sides of the triangle $p_1p_2p_3$ at q_1 and q_2 so that

$$\frac{p_2q_1}{p_3q_1} = \frac{P_2Q_1}{P_3Q_1}; \qquad \frac{p_1q_2}{p_3q_2} = \frac{P_1Q_2}{P_3Q_2}.$$

The intersection of the transversals p_1q_1 and p_2q_2 then gives a better approximation for the desired root. The assumption is that the points p_1, p_2, p_3 are not collinear.

Occasionally it can be advantageous for calculation of the complex roots of an equation to break up the equation into its real and imaginary parts, and calculate the roots of these two equations for the two variables. In place of the pair of linear equations above, we get the mapping equation of the region about the root, according to well-known theorems of function theory:[5]

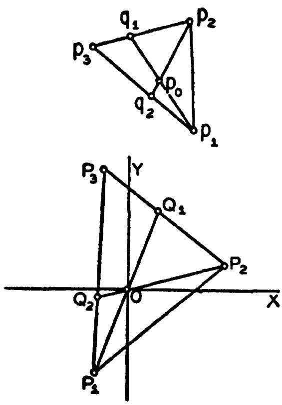


FIG. 69

$$(4) \qquad \begin{aligned} X &= a(x - x_0) - b(y - y_0); \\ Y &= b(x - x_0) + a(y - y_0), \end{aligned}$$

so that instead of the affine relation, we get a rotation through an angle φ and a scale change μ, where

$$(5) \quad \varphi = -\operatorname{arc\,tg}\frac{b}{a}, \qquad \mu = (a^2 + b^2)^{1/2}.$$

We therefore have a conformal mapping (cf. 21.9).

4. *Example*: For the determination of the exponent in the formula for longitudinal oscillations of an airplane at $v = 30$ m/sec. Deimler[6] gets the equation

$$\lambda^4 + 2\lambda^3 + 4.49\lambda^2 + 1.18\lambda + 0.7 = 0.$$

If we set $\lambda = x + iy$, then we get, by separation of the real and imaginary parts

$$X = x^4 - 6x^2y^2 + y^4 + 2x^3 - 6xy^2 + 4.49(x^2 - y^2)$$
$$+ 1.18x + 0.7,$$

$$Y = y(4x^3 - 4xy^2 + 6x^2 - 2y^2 + 8.98x + 1.18).$$

From this we obtain the following approximation values:

$$x_1 = -0.9, \quad y_1 = 1.7, \quad X_1 = -0.59, \quad Y_1 = -0.58,$$

$$x_2 = -0.9, \quad y_2 = 1.8, \quad X_2 = +0.2, \quad Y_2 = +0.41,$$

$$x_3 = -0.8, \quad y_3 = 1.8, \quad X_3 = +1.03, \quad Y_3 = -0.58.$$

These values are plotted in Fig. 70. The corresponding angle transversals are drawn in the XY plane and mapped into the xy plane. From this we read as a new approximation $x_4 = -0.893$, $y_4 = 1.769$.

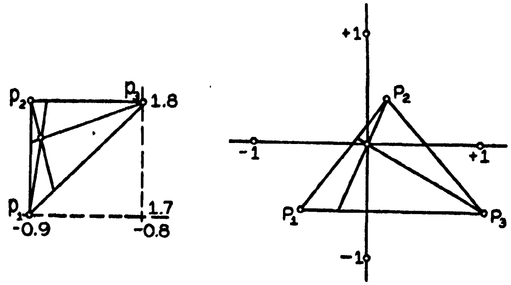


FIG. 70

If we substitute these values, we get $X_4 = 0.03$, $Y_4 = -0.06$. If this accuracy is not sufficient, we can continue in a similar manner.

Instead of calculating three points, since we are dealing with a conformal mapping, it is possible to work with two points P_1P_2 and construct $\Delta P_1P_2O \approx \Delta p_1p_2p$ we shall do in 21.10.

5. If the *method of iteration* is to be applied to the solution of an equation, then it is assumed that the equation may be put in the form

$$x = \varphi(x), \tag{6}$$

where $\varphi(x)$ is a function of x which changes slowly in the region of the root x_0. By iteration we understand the following: if we have found an approximate value x_1 for the root of the equation, then we can approximate $\varphi(x_1) \approx \varphi(x_0)$ because of the small variation of $\varphi(x)$, and

$$x_2 = \varphi(x_1) \tag{7}$$

will be a better approximation of the root. With x_2 we form $x_3 = \varphi(x_2)$ and continue until the root is found with sufficient accuracy. If in the method of false position mentioned in Sec. 1 we get the value lying on the convex side of the curve, say $\bar{x}_1$, then we can put the equation (1) in the form

$$x = \frac{x\bar{y} - \bar{x}y}{\bar{y} - y} = \frac{x\bar{y} - \bar{x}f(x)}{\bar{y} - f(x)},$$

i.e., we can consider the method of false position as an iteration process.

To provide us with a survey of this type of approximation, we form

$$x_0 - x_2 = \varphi(x_0) - \varphi(x_1) \tag{8}$$

by subtraction of the two equations above. If we are working with real values of x, then there are two possibilities for the approximation. First, the approximation for the root can be approached from one side. Then the approximation values form a monotonic increasing or decreasing sequence. This is the case if $\varphi(x)$ increases in the neighborhood of the root with increasing x. For then, if $x_0 > x_1$, $\varphi(x_0) > \varphi(x_1)$, $x_0 - x_2 > 0$, i.e., $x_0 > x_2$ lies on the same side of x_0 as x_1, and indeed on the side of the smaller x in this case. If, on the other hand, $\varphi(x)$ decreases with increasing x in the neighborhood of x, then, if $x_0 > x_1$, $\varphi(x_0) < \varphi(x_1)$, i.e., $x_0 - x_2 < 0$ or $x_0 < x_2$, x_2 lies on the opposite side of the root from x_1.

To make the convergence of the process completely clear to us, we form, from equations (6) and (7):

$$x - x_2 = \frac{\varphi(x) - \varphi(x_1)}{x - x_1}(x - x_1). \tag{9}$$

For continuous functions—and we consider only such functions here—the difference quotient is always smaller than the maximum value of the derivative in the interval. If the absolute value of this derivative in the observed interval has the maximum value m, then $| x - x_2 | \leqq m | x - x_1 |$, and further $| x - x_3 | < m | x - x_2 | \leqq m^2 | x - x_1 |$. In general

$$| x - x_n | \leqq m^{n-1} | x - x_1 |. \tag{10}$$

Therefore the iteration process converges if, in the interval used for approximation, the maximum absolute value of the derivative of $\varphi(x)$ is smaller than 1. Indeed, the smaller this value, the more rapid is the convergence. It is to be observed with this derivative that nothing has been assumed concerning the real nature of the roots, the approximation value, or the coefficients entering into them. The method of iteration therefore can be applied to the calculation of complex roots of such complex functions.

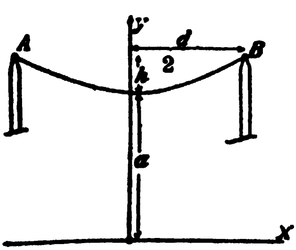


FIG. 71

A particular advantage of the iteration process is that an occasional error of calculation does not ruin the work, provided that the erroneously computed value lies within the region of convergence. Under such circumstances, it adds only to the amount of work, so that in spite of occasional errors in calculation, we arrive at the desired final value.[7]

6. *Example*: An inelastic chain is to be stretched between two points

at the same altitude, $d = 50$ m. apart. The chain sags $h = 5$ m. at the center. How long is the chain?

The equation of the chain curve is

$$y = a \cosh x/a, \text{ its length}$$

$$l = 2 \int_0^{d/2} \left(1 + \sinh^2 \frac{x}{a}\right)^{1/2} dx = 2a \int_0^{d/2} \cosh \frac{x}{a}\, d\left(\frac{x}{a}\right) = 2a \sinh \frac{d}{2a}.$$

For the determination of the parameter, we have the equation

$$a + h = a \cosh \frac{d}{2a}$$

$$= a\left[1 + \frac{1}{2}\left(\frac{d}{2a}\right)^2 + \frac{1}{24}\left(\frac{d}{2a}\right)^4 + \frac{1}{720}\left(\frac{d}{2a}\right)^6 + \cdots\right]$$

$$\frac{2h}{d}\frac{d}{2a} = \frac{1}{2}\left(\frac{d}{2a}\right)^2 + \frac{1}{24}\left(\frac{d}{2a}\right)^4 + \frac{1}{720}\left(\frac{d}{2a}\right)^6 \cdots$$

or, if we set $d/(2a) = z$,

$$\frac{2h}{d} z = \frac{1}{2} z^2 + \frac{1}{24} z^4 + \frac{1}{720} z^6 \cdots .$$

If we divide by z and rearrange terms, we have

$$z = +\frac{4h}{d} - \frac{z^3}{12} - \frac{z^5}{360} = \cdots = \varphi(z).$$

From this, z can be computed by the method of iteration, if the sagging is small. As a first approximation, we take

$$z_1 = +\frac{4h}{d} = 0.4.$$

From this it follows that

$$z_2 = 0.4 - 0.0053 - 0.0001 = 0.3946.$$

As a second approximation, we find

$$z_3 = 0.39485.$$

Since the value of $\varphi'(z) \approx -0.040$ is accurate to three decimals, and the error of the first approximation can be estimated at a maximum of 0.006 from the above data, the error of z_3 is

$$z_0 - z_3 < m^2(z - z_1) \approx 0.04^2 \times 0.006 \approx 0.0000096,$$

i.e., the value of z_3 is accurate to about a unit in the fifth decimal place. The next approximation value

$$z_4 = 0.394843$$

is accurate to the sixth place. From this we may calculate

$$a = \frac{d}{2z} = \frac{25}{0.394843} = 63.3160$$

and obtain the value of the length of the chain:

$$l = 126.632 \sinh (0.394843) \approx 51.18 \text{ m.}$$

Other problems on the catenary can be solved in similar fashion.

7. The *process of iteration* can be extended to *several equations with several unknowns,* provided that these can be put in suitable form. For general observations, we limit ourselves to two equations with two unknowns. The treatment of several equations with several unknowns can be inferred immediately. Both equations must be put in the form

$$x = \varphi(x, y), \qquad y = \psi(x, y) \tag{11}$$

if we want to use the iteration process, where φ and ψ vary only slightly in the neighborhood of the desired root. If a first approximation x_1, y_1 has been found in some way, perhaps graphically, better approximations can be obtained in the same way as above:

$$x_2 = \varphi(x_1, y_1), \qquad y_2 = \psi(x_1, y_1). \tag{12}$$

From these two approximations, a third can be obtained, etc.

To obtain information on the convergence of the process, we subtract (11) from (12):

$$x - x_2 = \varphi(x, y) - \varphi(x_1, y_1); \qquad y - y_2 = \psi(x, y) - \psi(x_1, y_1) \tag{12a}$$

or, if we rearrange the right hand side of the first equation,

$$\begin{aligned} x - x_2 = {} & \frac{\varphi(x, y) - \varphi(x_1, y)}{x - x_1}(x - x_1) \\ & + \frac{\varphi(x_1, y) - \varphi(x_1, y_1)}{y - y_1}(y - y_1). \end{aligned} \tag{13}$$

If, instead of the difference quotients, we insert the maxima of the absolute values of the partial derivatives $|\bar{\varphi}_x|$ and $|\bar{\varphi}_y|$ for the region considered, we obtain

$$x - x_2 | \leqq |\bar{\varphi}_x| \cdot |x - x_1| + |\bar{\varphi}_y| \cdot |y - y_1|, \tag{14}$$

and, in the same way,

$$|y - y_2| \leqq \bar{\psi}_x| \cdot |x - x_1| + |\bar{\psi}_y| \cdot |y - y_1|. \tag{14a}$$

From this it follows, by addition, that

$$
(15) \qquad \begin{aligned} |x - x_2| + |y - y_2| \leqq (|\bar{\varphi}_x| + |\bar{\psi}_x|) \cdot |x - x_1| \\ + (|\bar{\varphi}_y| + |\bar{\psi}_y|) \cdot |y - y_1|. \end{aligned}
$$

If we denote the maximum value of these four derivatives in the observed region by $m/2$, we get

$$
(15a) \qquad |x - x_2| + |y - y_2| \leqq m(|x - x_1| + |y - y_1|),
$$

and, in general, as in the case of one unknown,

$$
(16) \qquad |x - x_n| + |y - y_n| \leqq m^{n-1}(|x - x_1| + |y - y_1|).
$$

The method is therefore certainly convergent if the partial derivatives of the functions on the right side of the equations are all less that $1/l$, where l is the number of unknowns entering into the equations. This can be seen by an extension of the above considerations to l equations with l unknowns. If we apply this process to linear equations, we obtain, essentially, the method of Seidel[8] applied to the solution of the normal equations in the method of least squares.

8. *Example*: An example in which the above condition is not completely fulfilled, while the process nevertheless converges, although slowly, is taken from a work by *Ritz*[9] on the *Chladni figures* of a square plate.

For the six coefficients A_0 to A_5 there are homogeneous linear equations in which the parameter λ appears; equations which are independent of each other if the determinant is zero. This gives an equation of sixth order for the evaluation of the frequency λ, which has the form of the secular equation. These equations are:

$$
\begin{aligned}
0 &= (13.95 - \lambda) A_0 - 32.08 A_1 + 18.60 A_2 \\ &\quad + 32.08 A_3 - 37.20 A_4 + 18.60 A_5 \\
0 &= - 16.04 A_0 + (411.8 - \lambda) A_1 - 120.0 A_2 \\ &\quad - 133.6 A_3 + 166.80 A_4 + 140 A_5 \\
0 &= + 18.60 A_0 - 240.0 A_1 + (1686 - \lambda) A_2 \\ &\quad - 218 A_3 - 1134 A_4 + 330 A_5 \\
0 &= + 16.04 A_0 - 133.6 A_1 + 109.0 A_2 \\ &\quad + (2945 - \lambda) A_3 - 424 A_4 + 179 A_5 \\
0 &= - 18.60 A_0 + 166.8 A_1 - 567 A_2 \\ &\quad - 424 A_3 + (6303 - \lambda) A_4 + 1437 A_5 \\
0 &= + 18.60 A_0 + 280 A_1 + 330 A_2 \\ &\quad + 358 A_3 - 2874 A_4 + (13674 - \lambda) A_5 .
\end{aligned}
$$

Since the coefficients are calculated with a slide rule, the rule can also be used for carrying out the calculations. Only the ratios of the coefficients can be obtained from the above equations. Therefore A_0 is set equal to one, and a first approximation is taken for λ at 13.95. If this is substituted in the last five equations, the diagonal terms are predominant and, by neglecting the other terms, we get as a first approximation $A_1 = 0.04035$; $A_2 = -0.01112$; $A_3 = -0.00547$; $A_4 = 0.00296$; $A_5 = -0.00136$. These, substituted in the first equation, give as a correction for λ:

$$\Delta\lambda = -32.08\, A_1 + 18.60\, A_2 + 32.08\, A_3 - 37.20\, A_4 + 18.60\, A_5$$

$$= -1.293 - 0.207 - 0.175 - 0.110 - 0.025 = -1.810.$$

Therefore $\lambda_2 = 12.14$. This value is substituted again on the right side of the last five equations, which are put in the form

$$(411.8 - \lambda)\, A_1 = 16.04 + 120\, A_2 + 133.6\, A_3 - 166.8\, A_4 - 140\, A_5$$

$$(1686 - \lambda)\, A_2 = -18.60 + 240\, A_1 + 218.0\, A_3 + 1134\, A_4 - 330\, A_5$$

. . . .

From this, a new set of values $A_1 \cdots A_5$ are obtained; with these a new value of λ is obtained, and so forth. The sequences of approximation values are given below:

TABLE III.

	I	II	III	IV	V	VI	VII
A_1	+0.04035	+0.0342	+0.0384	+0.0372	+0.0379	+0.03765	+0.03790
A_2	−0.01112	−0.00375	−0.00620	−0.00484	−0.00539	−0.00480	−0.00491
A_3	−0.00547	−0.00272	−0.00367	−0.00320	−0.00340	−0.00329	−0.00335
A_4	+0.00296	+0.00020	+0.00127	+0.00071	+0.00095	+0.00083	+0.00091
A_5	−0.00136	−0.00115	−0.00186	−0.00164	−0.00178	−0.00172	−0.00176
λ	12.14	12.66	12.40	12.51	12.45	12.49	12.47

The calculation of harmonics can be performed in a way similar to the above. We set $A_1 = 1$, take as a first approximation $\lambda = 412$, and calculate the coefficients from the first and third to the sixth equation. The convergence is very slow because of the magnitude of the partial derivatives. If no parameter enters into the equations, we can make the coefficients small in comparison with the diagonal terms

by multiplication of one or another equation with a suitable factor, and subtraction from the other equations.

9. In the secular equations considered in the above example, it is also possible to remove the terms not on the diagonal by a transformation due to Jacobi,[10] thereby improving the convergence of the process. If the coefficients of the equation are designated by a double index, the first digit of which gives the number of the equation, the second the number of the unknown, we have

$$\begin{aligned} g_1 &: a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \cdots + a_1 = 0, \\ g_2 &: a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \cdots + a_2 = 0, \\ g_3 &: a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + \cdots + a_3 = 0 \text{ etc.,} \end{aligned} \tag{17}$$

where all coefficients on the diagonal may be large compared to the others, except that either or both a_{12} or a_{21} may be large and thereby impair the convergence of the interation process. If some other coefficients are involved, the calculations follow in the same way. We set

$$\begin{aligned} \cos 2\Delta \cdot x_1 &= \cos(\alpha + \Delta)\xi_1 + \sin(\alpha - \Delta)\xi_2 , \\ \cos 2\Delta \cdot x_2 &= \sin(\alpha + \Delta)\xi_1 - \cos(\alpha - \Delta)\xi_2 , \end{aligned} \tag{18}$$

where α and Δ are determined from the equations

$$\rho \cos 2\alpha = a_{11} - a_{22} ; \qquad \rho \sin 2\alpha = a_{12} + a_{21} ; \qquad \rho \sin 2\Delta = a_{21} - a_{12} . \tag{19}$$

If we put the new equations

$$\gamma_1 = g_1 \cos(\alpha - \Delta) + g_2 \sin(\alpha - \Delta); \qquad \gamma_2 = g_1 \sin(\alpha + \Delta) - g_2 \cos(\alpha + \Delta) \tag{20}$$

in place of the first two equations g_1 and g_2 in (17), then the coefficients α_{12} and α_{21} vanish in these equations, as can easily be determined by substitution. We then get the system of equations

$$\begin{aligned} \gamma_1 &: \alpha_{11}\xi_1 \qquad\qquad + \alpha_{13}x_3 + \cdots + \alpha_1 = 0, \\ \gamma_2 &: \qquad + \alpha_{22}\xi_2 + \alpha_{23}x_3 + \cdots + \alpha_2 = 0, \\ g_3 &: \alpha_{31}\xi_1 + \alpha_{32}\xi_2 + a_{33}x_3 + \cdots + a_3 = 0, \end{aligned} \tag{21}$$

where

$$\alpha_{11} = \frac{a_{11} + a_{22}}{2} + \frac{\rho}{2} \cos 2\Delta, \qquad \alpha_{22} = \frac{a_{11} + a_{22}}{2} - \frac{\rho}{2} \cos 2\Delta,$$

$$(22) \quad \begin{aligned} \alpha_{1n} &= a_{1n} \cos(\alpha - \Delta) + a_{2n} \sin(\alpha - \Delta); \\ \alpha_{2n} &= a_{1n} \sin(\alpha + \Delta) - a_{2n} \cos(\alpha + \Delta); \\ \alpha_{n1} \cos 2\Delta &= a_{n1} \cos(\alpha + \Delta) + a_{n2} \sin(\alpha + \Delta); \\ \alpha_{n2} \cos 2\Delta &= a_{n1} \sin(\alpha - \Delta) - a_{n2} \cos(\alpha - \Delta), \end{aligned}$$

as may be seen by substitution.

For a check on the calculation, we can use the fact that

$$(23) \qquad \alpha_{11} + \alpha_{22} = a_{11} + a_{22}\,; \qquad \alpha_{1n}\alpha_{n1} + \alpha_{2n}\alpha_{n2} = a_{1n}a_{n1} + a_{n2}a_{n2}\,.$$

10. In the calculation of examples of Sec. 8, the coefficients $a_{21} = -120$ and $a_{12} = -240$ are at first disagreeably large. To remove these coefficients, we set

$$(23\text{a}) \qquad \rho \cos 2\alpha = 1274.2, \qquad \rho \sin 2\alpha = -360, \qquad \rho \sin 2\Delta = 120.$$

From this follows

$$\alpha = 172°6.7', \qquad \Delta = 2°36'.0.$$

By a continuation of the computations given above, we then get the system of equations

$$\begin{aligned}
&(13.95-\lambda)\, A_0 - 28.27\, \overline{A_1} - 21.55\, \overline{A_2} + 32.08\, A_3 - 37.20\, A_4 + 18.60\, A_5 = 0 \\
\Gamma_1 :\ &14.26\, A_0 + (389.6-\lambda)\, \overline{A_1} - 153.13\, A_3 + 61.57\, A_4 + 169.83\, A_5 = 0 \\
\Gamma_2 :\ &-21.21\, A_0 + (1708.2-\lambda)\, \overline{A_2} + 190.04\, A_3 + 1145.41\, A_4 - 299\, A_5 = 0 \\
G_3 :\ &+16.04\, A_0 - 111.99\, \overline{A_1} - 121.22\, \overline{A_2} + (2945-\lambda)\, A_3 - 424\, A_4 + 179\, A_5 = 0 \\
G_4 :\ &-18.60\, A_0 + 61.05\, \overline{A_1} + 581.71\, \overline{A_2} - 424\, A_3 + (6303-\lambda)\, A_4 - 1437\, A_5 = 0 \\
G_5 :\ &+18.60\, A_0 + 336.77\, \overline{A_1} - 303.66\, \overline{A_2} + 358\, A_3 - 2874\, A_4 + (13674-\lambda) A_5 = 0.
\end{aligned}$$

In this, the terms $a_{45} = -1437$ and $a_{54} = -2874$ are especially annoying. To remove these, we must set

$$\rho \cos 2\alpha = 7371, \qquad \rho \sin 2\alpha = -4311, \qquad \rho \sin 2\Delta = 1437,$$

and we get $\alpha = 164°50.35'$, $\Delta = 4°50.62'$. The new equations are then

$$(13.95-\lambda)\, A_0 - 28.27\, \overline{A_1} - 21.55\, \overline{A_2} + 32.08\, A_3 - 29\, \overline{A_4} - 25.32\, \overline{A_5} = 0$$

$$-14.26\, A_0 + (389.6-\lambda)\, \overline{A_1} - 153.13\, A_3 + 117.63\, \overline{A_4} - 158.31\, \overline{A_5} = 0$$

$$-21.21\, A_0 \quad (1708.2-\lambda)\, \overline{A_2} + 190.04\, A_3 + 988.13\, \overline{A_4} + 506.63\, \overline{A_5} = 0$$

$$+16.04\, A_0 - 111.99\, \overline{A_1} - 121.22\, \overline{A_2} + (2945-\lambda)\, A_3 - 342.07\, \overline{A_4} + 255.69\, \overline{A_5} = 0$$

$$-14.97\, A_0 + 120.37\, \overline{A_1} + 517.93\, \overline{A_2} - 353.03\, A_3 + (5779.8-\lambda)\, \overline{A_4} = 0$$

$$-23.84\, A_0 - 295.54\, \overline{A_1} + 484.35\, \overline{A_2} - 481.45\, A_3 + (14197.2-\lambda)\, \overline{A_5} = 0.$$

The iteration process then gives

	I	II	III	IV	V	VI
$\overline{A_1}$	+0.0380	+0.0355	+0.0373	+0.0370	+0.03715	+0.03715
$\overline{A_2}$	+0.01251	+0.01111	+0.01210	+0.01192	+0.01200	+0.01198
A_3	−0.00547	−0.00305	−0.00345	−0.00330	−0.00334	−0.00332
$\overline{A_4}$	+0.00259	+0.00034	+0.00067	+0.00052	+0.00055	+0.000540
$\overline{A_5}$	+0.00168	+0.00186	+0.00194	+0.001935	+0.001935	+0.001933
λ	12.35	12.55	12.46	12.47	12.47	12.47

Further variations do not occur. If, from these values, we calculate the original values A according to the equations (18), we get

$$A_1 = +0.03779, \qquad A_2 = -0.00518, \qquad A_3 = -0.00332,$$
$$A_4 = +0.000866, \qquad A_5 = -0.001742.$$

If the iteration process is repeated with these values in the initial equations of the example in Sec. 8, we get the same results. When we begin with the iteration process, the rearrangement of the equations must be left to the experience of the computer. If it is a question of determining the collected values λ, it can be advantageous to reduce all members off the diagonal to zero by the Jacobi transformation.

11. If we want to calculate the real roots of a real function by *Newton's method of approximation,* we replace the curve representing the function, not with a chord as was the case with the method of false position, but with a tangent. The equation of the tangent at a point with abscissa x_1 is given by

$$\frac{y - f(x_1)}{x - x_1} = f'(x_1),$$

and its x intercept is

$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)}, \tag{24}$$

i.e., the correction which is added to x_1 to get x_2 is

$$\Delta x_1 = - \frac{f(x_1)}{f'(x_1)}, \tag{25}$$

a value which is also obtained if $f(x)$ is developed about zero in a *Taylor series,* and the expansion is terminated with the linear term. If we calculate the value of the function for x_2 and if this still differs greatly from zero, we must repeat the calculation once more for x_2.

To find out the best way to begin this approximation method, we first assume that we have a simple root, that therefore $f'(x_0) \neq 0$, that there is no point of inflection in the interval, i.e., that $f''(x) \neq 0$ in that interval. As may be read from Fig. 72, we then get a sequence of values approaching the root from one side, in Fig. 72 from $x_1 < x_0$. If, on the other hand, we

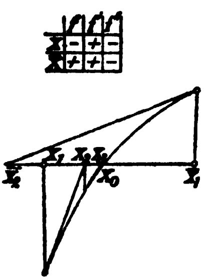


FIG. 72

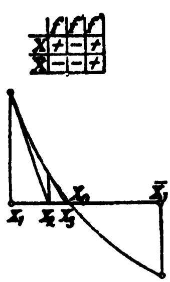


FIG. 73

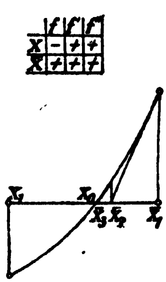


FIG. 74

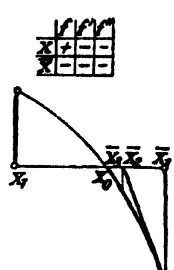


FIG. 75

start out from the other value $\bar{x}_1$, then we get a new value $\bar{x}_2$ which lies on the other side of x_1, so that we get, under certain circumstances, no sequence of values approaching the root. It is therefore advisable to begin with a value which gives a monotonic increasing or decreasing sequence of roots. In addition to the case represented in Fig. 72, there are also the

three cases of Fig. 73 to Fig. 75. From these may be seen that the function and its second derivative have the same sign at a useful approximation point.

To determine the degree of approximation attained in each step, we observe that the Newton method is only a special case of the iteration process. If x satisfies the equation $f(x) = 0$, it also satisfies the equation

$$x = x - \frac{f(x)}{f'(x)} = \varphi(x), \tag{26}$$

if $f'(x) \neq 0$. The iteration process can be used on this equation and the Newton formula given above is obtained:

$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)} = \varphi(x_1).$$

By consideration of Sec. 5 we obtain a better approximation value, if the maximum absolute value of the first derivative of $\varphi(x)$ is smaller than one in the approximation interval, i.e., if

$$\mid \varphi'(x) \mid = \left| \frac{f(x) \cdot f''(x)}{(f'(x))^2} \right| < 1. \tag{27}$$

From the considerations in Sec. 5 of this article, the approximation approaches the value of the root from one side if $\varphi' > 0$, i.e., in agreement with Fig. 72-75, we must have a first approximation value with real roots for which

$$f(x) \cdot f''(x_1) > 0.$$

The approximation values then lie on the convex side of the curve. Since the method of false position gives a better approximation value lying on the concave side, we can combine the two methods[11] and obtain simultaneously new approximation values for the upper and lower limits of the interval. On the other hand, in the application of the method of false position, either the second limit value must remain unchanged, or tests must be made for the improvement of this value. To estimate the rapidity of the convergence of the approximation, we form

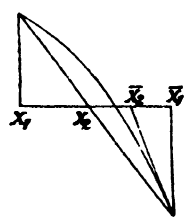


FIG. 76

$$(x_0 - x_2) = \varphi(x_0) - \varphi(x_1).$$

If we develop $\varphi(x_1)$ about the root x_0 in a Taylor series, and neglect the powers of $(x_1 - x_0)$ which are higher than the second (which is permissible because of the smallness of $(x_1 - x_0)$), then we obtain, if we observe that

$$\varphi'(x_0) = \frac{f(x_0)f''(x_0)}{[f'(x_0)]^2} = 0,$$

$$x_0 - x_2 = \varphi(x_0) - \varphi(x_0) - \frac{1}{2}\varphi''(x_0)(x_1 - x_0)^2 = -\frac{1}{2}\frac{f''(x_0)}{f'(x_0)}(x_1 - x_0)^2,$$

$$(28) \qquad | x_2 - x_0 | \leqq m(x_1 - x_0)^2 = \frac{[m(x_1 - x_0)]^2}{m}.$$

In the formation of the second derivative, it is again to be observed that $f(x_0) = 0$, also that the absolute value of $-f''(x_0)/2f'(x_0)$ is represented by m. In this way we find that

$$(29) \qquad \begin{aligned} | x_3 - x_0 | &\leqq m(x_2 - x_0)^2 \leqq \frac{1}{m}[m(x_1 - x_0)]^4 \\ | x_4 - x_0 | &\leqq \frac{1}{m}[m(x_1 - x_0)^8] \text{ etc.} \end{aligned}$$

The errors increase in proportion to the 2nd, 4th, $\cdots$ powers of $m(x_1 - x_0)$. If we select x_1 near enough to the root x_0 that this expression is smaller than 1, the process converges extremely rapidly.

For carrying out the computation, it is often inconvenient to calculate $f'(x)$, particularly if $f(x)$ is given in tables; then we take two adjacent tabular values and divide the difference of the two Δx, i.e., replace the derivative by the difference quotient $\Delta f(x)/\Delta x$.

Since here also nothing has been assumed concerning the real nature of the values used, the observations on convergence are also valid for complex roots. It is only necessary that

$$(30) \qquad \left| \frac{1}{2}\frac{f''(x_0)}{f'(x_0)} \right| \cdot | x_0 - x_1 | < 1$$

in the interval considered, e.g., 21.11. The approximation values converge to the desired value of the root on a spiral curve.

12. *Example*: In the determination of the characteristic vibrations of an elastic bar and also in other problems, there occurs the equation

$$f(x) = \text{tg } x + \tanh x = 0, \qquad f'(x) = \frac{1}{\cos^2 x} + \frac{1}{\cosh^2 x}.$$

From this it follows that

$$f''(x) = \frac{2 \text{ tg } x}{\cos^2 x} - \frac{2 \tanh x}{\cosh^2 x}.$$

Therefore

$$m = \left| -\frac{1}{2}\frac{f''(x)}{f'(x)} \right| = \left| \frac{-\text{tg } x \cosh^2 x + \tanh x \cos^2 x}{\cosh^2 x + \cos^2 x} \right|.$$

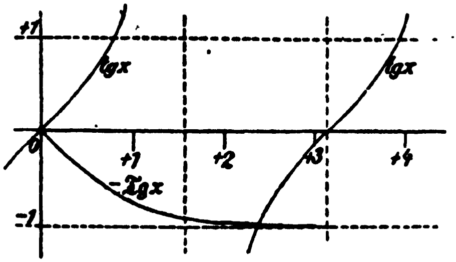


FIG. 77

From Fig. 77, it follows that the first root of $f(x)$ lies in the vicinity of $3\pi/4$. Therefore,

$$m \approx \left| \frac{5.32^2 + 1/2 \times 0.98}{5.32^2 + 1/2} \right| \approx 1.$$

If therefore $|x_1 - x_0| < 1$, then the process is convergent. If, for example, we take the value $x_1 = 2.36$, which is roughly $3\pi/4$, then this is certainly the case. Since tables exist for tg x and tanh x, we replace the derivative by the difference quotient as mentioned above. This gives, if we use the tables of Hayashi[3]

x	x^0	tg x	tanh x	Δ_1	Δ_2
2.36	135° 13′ 4.94″	−0.9924	0.9823	0.00199	0.00003
2.3650	135° 30′ 16.27″	−0.98254	0.98250	0.00204	0.00004

$f(x)$	$\Delta f(x)$	$-\dfrac{f(x)\cdot\Delta x}{\Delta f(x)}$
−0.0101	0.00202	$\dfrac{1010\cdot 0.001}{202} = 0.0050$
−0.00004	0.00208	$\dfrac{4\cdot 0.001}{208} = 0.000019$

Therefore $x_3 = 2.365019$. We have begun here on the correct side, since $f(x_1)$ as well as $f''(x_1)$ are negative, i.e., they have the same

sign. The approximation values therefore form a monotonic increasing sequence. The error of x_1 is about $\Delta x_1 \approx 0.005$. The error of x_3 is therefore

$$\Delta x_3 = m^3 \Delta_1^4 x \approx 0.0000000007.$$

since $m \approx 1$. The places written out are then correct, if an inaccuracy does not enter into the last figure because of rounding off and interpolation errors. There is no point in further calculation, since the table is only a five place one.

For practical calculations, it is sufficient in most cases not to calculate $f'(x_n)$ each time, but to use $f'(x_0)$ for all cases. Tangents are then not used for the representation of the curve, but instead, lines parallel to the first tangent. The number of steps of approximation is then larger, but the work of calculation is appreciably less.

13. As mentioned above, the Newton method can be considered in the following way: we develop the function $y = f(x)$, whose roots are to be determined, in a Taylor series about a neighboring point x_1 of the root x_0:

$$(31)\ f(x_0) = f(x_1) + \frac{f'(x_1)}{1!} \Delta x_1 + \frac{f''(x_1)}{2!} (\Delta x_1)^2 + \frac{f'''(x_1)}{3!} (\Delta x_1)^3 \cdots = 0.$$

We terminate this series with the linear term in Δx_1 and calculate the correction from it. Now it is possible to consider additional terms in the development, i.e., to replace the curve, not by a tangent, but by a curve of higher order, which, at the point $x = x_1$, has the higher number of derivatives in common with the given curve.

We can consider the value of Δx_1 so determined as expanded in a power series in $f(x_1)$:

$$(32) \qquad \Delta x_1 = af(x_1) + b(f(x_1))^2 + c(f(x_1))^3 \cdots .$$

This will lead to a more convenient calculation of the approximation value. If (32) is substituted in (31) and the method of undetermined coefficients is used, i.e., if the factors of similar exponents of $f(x_1)$ are set equal to zero, we get

$$0 = f(x_1) + [af(x_1) + b(f(x_1))^2 + c(f(x_1))^3 \cdots] f'(x_1)$$

$$+ [af(x_1) + b(f(x))^2 \cdots]^2 \frac{f''(x_1)}{2} + [af(x_1) \cdots]^3 \frac{f''(x_1)}{6_1},$$

$$(32a) \qquad 1 + af'(x_1) = 0, \qquad a = -\frac{1}{f'(x_1)}.$$

$$bf'(x_1) + \frac{1}{2}a^2f''(x_1) = 0, \qquad b = -\frac{1}{2}\frac{f''(x_1)}{(f'(x_1))^3},$$

$$cf'(x_1) + abf''(x_1) + \frac{1}{6}a^3f'''(x_1) = 0, \; c = -\frac{(f''(x_1))^2}{2(f'(x_1))^5}$$

$$+ \frac{f'''(x_1)}{6(f'(x_1))^4}.$$

If we omit the argument, for brevity, then

$$x_2 = x_1 - \frac{f}{f'} - \frac{f''}{2f'^3}\cdot f^2 - \left(\frac{f''^2}{f'} - \frac{f'''}{3}\right)\frac{f^3}{2f'^4} \cdots . \tag{33}$$

Of course such a formula is of use only if the derivatives can be easily calculated, or can be obtained from tables.

14. *Example*: If the positive root of the equation

$$f(x) = \sin x - \frac{x}{2} = 0 \tag{33a}$$

is to be determined, an approximation $x_1 = 1.9$ is easily found by graphing. From this it results, if we take $f(x_1)$ to seven decimals, and the derivatives from Hayashi's tables to five decimals,

$$\begin{aligned} f(1.9) &\doteq -0.0037000, \\ f'(1.9) &= -0.82329, \\ f''(1.9) &= -0.94630 \end{aligned} \tag{33b}$$

and if a second approximation is sufficient,

$$\begin{aligned} x_2 &= 1.9 - \frac{0.0037000}{0.82329} - \frac{0.94630}{2(0.82329)^3}(0.0037000)^2 \\ &= 1.9 - 0.0044942 - 0.0000116 = 1.8954942. \end{aligned} \tag{33c}$$

The term with the third power f^3 gives a 5 in the eighth place after the decimal point, so that we can assume that the calculated value is correct to six decimal places.

15. Just as with the iteration process, the Newton method can be extended to equations with several unknowns. The Newton method consists essentially of a Taylor series expansion about a point in the neighborhood of the desired root. This series is broken off with the linear term.

The same can be done with several, or n variables, and there results for n arbitrary equations with n unknowns a system of n linear equations. The procedure is the same as employed in the error calculation of linear error equations. If we have two equations $f(x, y) = 0$ and $g(x, y) = 0$, then we set for the approximation values x_1, y_1:

$$\begin{aligned} f(x, y) &\approx f(x_1, y_1) + f_x(x_1, y_1)\cdot\Delta x_1 + f_y(x_1, y_1)\Delta y_1 = 0, \\ g(x, y) &\approx g(x_1, y_1) + g_x(x_1, y_1)\cdot\Delta x_1 + g_y(x_1, y_1)\Delta y_1 = 0. \end{aligned} \tag{34}$$

From these equations new approximation values x_2, y_2 are calculated, provided that the function determinant is $D = f_x g_y - f_y g_x \neq 0$. These new approximations are

$$x_2 = x_1 - \frac{fg_y - gf_y}{D}; \qquad y_2 = y_1 + \frac{fg_x - gf_x}{D}. \tag{35}$$

But this is just the same as if the iteration process had been applied to the equations

$$x = x - \frac{fg_y - gf_y}{D} = \varphi(x, y), \quad y = y + \frac{fg_x - gf_x}{D} = \psi(x, y). \tag{36}$$

If x_0, y_0 are the desired roots, we can set

$$x_0 - x_2 = \varphi(x_0, y_0) - \varphi(x_1, y_1); \quad y_0 - y_2 = \psi(x_0, y_0) - \psi(x_1, y_1), \tag{36a}$$

and can expand about the point in a Taylor series:

$$\begin{aligned} \varphi(x_1, y_1) &= \varphi(x_0, y_0) + \varphi_x(x_0, y_0)(x_1 - x_0) + \varphi_y(x_0, y_0)(y_1 - y_0) \\ &\quad + \frac{1}{2}[\varphi_{xx}(x_1 - x_0)^2 + 2\varphi_{xy}(x_1 - x_0)(y_1 - y_0) + \varphi_{yy}(y_1 - y_0)^2], \\ \psi(x_1, y_1) &= \psi(x_0, y_0) + \psi_x(x_0, y_0)(x_1 - x_0) + \psi_y(x_0, y_0)(y_1 - y_0) \\ &\quad + \frac{1}{2}[\psi_{xx}(x_1 - x_0)^2 + 2\psi_{xy}(x_1 - x_0)(y_1 - y_0) + \psi_{yy}(y_1 - y_0)^2]. \end{aligned} \tag{36b}$$

Since now $f(x_0, y_0) = 0$ and $g(x_0, y_0) = 0$, then

$$\begin{aligned} \varphi_x(x_0, y_0) &= 1 - \frac{f_x g_y - g_x f_y}{D} = 0, \qquad \psi_x(x_0, y_0) = \frac{f_x g_x - g_x f_x}{D} = 0, \\ \varphi_y(x_0, y_0) &= \frac{f_y g_y - g_y f_y}{D} = 0, \qquad \psi_y(x_0, y_0) = 1 + \frac{f_y g_x - g_y f_x}{D} = 0. \end{aligned} \tag{36c}$$

If the largest of the absolute values of the six second derivatives in the approximation interval is denoted by m, then

$$x_0 - x_2 = \varphi(x_0, y_0) - \varphi(x_1, y_1)$$

$$\leqq \frac{m}{2}[|x_0 - x_1|^2 + 2|x_0 - x_1| \cdot |y_0 - y_1| + |y_3 - y_1|^2]$$

(36d)

$$\leqq \frac{m}{2}[|x_0 - x_1| + |y_0 - y_1|]^2$$

$$y_0 - y_2 \leqq \frac{m}{2}[|x_0 - x_1| + |y_0 - y_1|]^2.$$

As in the case of one unknown,

$$|x_0 - x_2| + |y_0 - y_2| \leqq \frac{[m|x_0 - x_1| + m|y_0 - y_1|]^2}{m},$$

and in general

(37) $$|x_0 - x_n| + |y_0 - y_n| \leqq \frac{1}{m}[m|x_0 - x_1| + m|y_0 - y_1|]^{2n-1}.$$

Therefore here also we have an extremely rapid convergence to the value of the root, if only the first approximation value is chosen sufficiently close to the root. It is usually difficult to get these first approximation values. With two equations in two variables, for example, we can find it by approximate sketches of the curves $f(x, y) = 0$ and $g(x, y) = 0$.

16. *Example*: An elastic steel wire ($E = 19{,}000$ kg/mm.,^{2} $\gamma = 0.0078$ kg per meter length, and 1 mm^2 cross section) is stretched between two supports which are $d = 2w = 100$ m apart. The wire has a sag of $h = 2.5$ m. What is its length $2L$ and the horizontal tension H?

If the parameter of this more general catenary is $a = H/\gamma$ then the following equations hold:[12]

$$w = \frac{d}{2} = a \ln\left[\frac{L}{a} + \left(1 + \frac{L^2}{a^2}\right)^{1/2}\right] + \frac{a\gamma}{E} L,$$

(37a)

$$h = a\left[\left(1 + \frac{L^2}{a^2}\right)^{1/2} - 1\right] + \frac{\gamma}{2E} L^2.$$

Approximation values are obtained if the chain is first assumed to be inelastic; by the example in Sec. 3, we then have

$$z = \frac{d}{2a} = \frac{w}{a}; \qquad z = \frac{2h}{w} - \frac{z^3}{12} \cdots,$$

from which results $z = 0.09992$, and from this,

$$a_1 = 500.40, \qquad L_1 = 50.085.$$

If we now develop the above equations in Taylor series, then

$$w - \left[a_1 \ln\left(\frac{L_1}{a_1} + \left(1 + \frac{L_1^2}{a_1^2}\right)^{1/2}\right) + \frac{a_1\gamma}{E} L_1\right]$$

$$= \Delta a\left[\ln\left(\frac{L_1}{a_1} + \left(1 + \frac{L_1^2}{a_1^2}\right)^{1/2}\right) - \frac{L_1}{a_1(1 + L_1^2/a_1^2)^{1/2}} + \frac{\gamma}{E} L_1\right]$$

$$+ \Delta L\left[\frac{1}{(1 + L_1^2/a_1^2)^{1/2}} + \frac{a_1\gamma}{E}\right]$$

$$h - \left[a_1\left(\left(1 + \frac{L_1^2}{a_1^2}\right)^{1/2} - 1\right) + \frac{\gamma}{2E} L_1^2\right] = \Delta a\left[\frac{1 - (1 + L_1^2/a_1^2)^{1/2}}{(1 + L_1^2/a_1^2)^{1/2}}\right]$$

$$+ \Delta L\left[\frac{L}{a_1(1 + L_1^2/a_1^2)^{1/2}} + \frac{\gamma L_1}{E}\right].$$

If we substitute the approximation values calculated above, it follows that

$$-0.0148 = \Delta a \times 0.000358 + \Delta L \times 0.995$$

$$-0.00250 = -\Delta a \times 0.004975 + \Delta L \times 0.996.$$

From these equations, calculation yields

$$\Delta L = -0.015, \qquad \Delta a = 0.204.$$

Therefore the better approximation

$$a_2 = 500.60, \qquad L_2 = 50.07$$

results. If the accuracy achieved is not sufficient, we must repeat the calculation with these values. With regard to the value of L we must observe that the elastic wire undergoes an elongation which is given by

$$\Lambda = \frac{La\gamma}{2E}\left(1 + \frac{L^2}{a^2}\right)^{1/2} + \frac{\gamma a^2}{2E} \ln\left[\frac{L}{a} + \left(1 + \frac{L^2}{a^2}\right)^{1/2}\right] = 0.01004,$$

so that the real length of the wire is $L = 50.08$ m. The horizontal tension is then

$$H = a\gamma = 500.6 \times 0.0078 \approx 4 \text{ kg}.$$

The wire can then be stretched much more.

NOTES

1. For proof, see Weber, *Lehrbuch der Algebra* I, Art. 118. Seliwanoff, *Lehrbuch der Differenzenrechnung*, Ch. II. No. 17 (Leipzig, 1904).

2. This method is frequently used in business arithmetic. For example, for the determination of the effective interest on a loan from tables for the cash value of a bond; cf. Loewy, *Mathematik des Geld- und Zahlungsverkehr* (Leipzig, 1920), p. 162, 176, 184. Occasionally a generalization of the above rule is also used, the so-called mean power; cf. *Goldziher, Z. f. angew. Math. u. Mech.* 7 (1927), p. 323.

3. K. Hayashi, *Fünfstellige Tafeln der Kreis- und Hyperbelfunktionen sowie der Funktionen e^x and e^{-x}* (Leipzig, 1921).

4. Scheffers, *Sitzungberichte der Berliner Math. Gesellschaft* (1916), pp. 29-34.

5. Bieberbach, *Lehrbuch der Funktionentheorie I* (Leipzig, 1923), p. 35.

6. Deimler, *Z. f. Motorluftschiffahrt und Flugtechnik* I (1910).

7. A generalization is found in Montessus, *Méthode générale de détermination des racines des équations numériques* (Louvain, 1910).

8. Seidel, *Münchner Abhandlungen*, **11** (1874), Part 3, pp. 81-108.

9. Ritz, *Annalen der Physik* **28** (1909), p. 763.

10. Jacobi, *Schuhmachers astronomische Nachrichten*, vol. 22, No. 523; Werke II, pp. 467-478.

11. Dandelin, *Mém. de l'Acad. Roy. de Bruxelles*, **3** (1826), p. 30.

12. Skrobanek, *Z. f. angew. Math. u. Mech.* II (1922), p. 472.

19. The Roots of Rational Integral Functions.

1. If we compare the development of the rational integral function as it was carried out in 7(8),

$$(1)\qquad g(x) = a_0' + a_1''(x - x_1) + a_2'''(x - x_1)^2 + \cdots + a_n(x - x_1)^n,$$

with the Taylor expansion

$$(1a)\qquad g(x) = g(x_1) + g'(x_1)(x - x_1) + \frac{g''(x_1)}{2!}(x - x_1)^2 + \cdots + \frac{g^{(n)}(x_1)}{n!}(x - x_1)^n,$$

it will be found that

$$(2)\qquad g(x_1) = a_0';\ [x_1x_1] = g'(x_1) = a_1'';\ 2![x_1x_1x_1] = 2\cdot g''(x_1) = a_2''' \cdots \quad n![\underbrace{x_1 \cdots x_1}_{n}] = n!a_n = g^{(n)}(x_1).$$

This also follows directly from the theorems on the divided differences with repeated argument (11.1). This fact is of importance for the approximate calculation of *the roots of a rational integral function by Newton's method.*

If an approximation value x_1 has been found for the value of the root, then, by means of the Horner scheme, the function is developed in powers

of $u = x - x_1$. Such a scheme needs to be used only twice to find a_0' and a_1''. Consequently, a new approximation value is obtained:

$$x_2 = x_1 - \frac{a_0'}{a_1''}. \tag{3}$$

With the new approximation value, the development in powers of $x - x_2$ is carried out, etc. The disadvantage of this type of calculation is that as one goes further, the numbers must be multiplied by an ever increasing number of digits.

Therefore it is better to carry out the development in powers of $u = x - x_1$ all the way through, and develop this function of u in powers of $v = u - u_2$, where u_2 is the correction to be added to x_2. Then this development is also carried out completely, etc. Then one has only to multiply[1] with numbers of fewer digits.

2. *Example*: As an example, we choose

$$x^3 - 30x^2 + 2358.7 = 0.$$

This equation has the approximate root $x_1 = 11$. We make a development about this point by the Horner scheme:

1	-30	0	2358.7
	11	-209	-2299
1	-19	-209	$\boxed{+59.7 = g(x_1)}$
	$+11$	-88	
1	-8	$\boxed{-297 = g'(x_1)}$	
	$+11$		
1	$\boxed{+3 = \tfrac{1}{2}g''(x_1).}$		

$g(x_1)$ and $g''(x_1)$ have the same signs. We therefore approach the value of the root from the correct side by a sequence of monotonically increasing approximation values. As a correction we obtain

$$u_1 = \Delta x = \frac{59.7}{297} = 0.201.$$

The error of this value is, according to 18(21),

$$F_2 \approx \frac{1}{2}\left|\frac{f''(x_1)}{f'(x_1)}\right|(x - x_1)^2 = \frac{1}{2}\frac{6}{297}\,0.2^2 \approx 0.0004.$$

To calculate further, we now start out from the function

$$u^3 + 3u^2 - 297u + 59.7 = 0$$

and develop about $u_2 = 0.20$, neglecting the last place:

1	3	−297	+59.7
	0.2	0.64	−59.272
1	3.2	−296.36	+0.428 = $g(u_1)$
	0.2	0.68	
1	3.4	−295.68 = $g'(u_1)$.	

If the constant term of the given equation is accurate only to a half unit of the last place, there is no point in carrying the calculation further; we therefore terminate the scheme. We have

$$v_3 = \Delta u_2 = \frac{0.428}{295.68} = 0.00145.$$

Therefore $x = 11.2015$. If the calculations had been carried out more accurately, the error of the third approximation would have been

$$F_3 = m^3(x - x_1)^4 \approx 0.01^3 \times 0.2^4 \approx 0.0000000002.$$

3. It frequently happens that several or all the *coefficients* a_n, a_{n-1}, $\cdots$, a_0 of the function are taken from experimental data, so that they possess *inaccuracies* $-\Delta_n$, $\cdots$, $-\Delta_0$. Consequently, the root x_0 which is evaluated from the equation

$$a_nx^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0 = 0, \tag{4}$$

has an error $-E$; therefore the correct value $x_0 + E$ must satisfy the equation

$$\begin{aligned}(a_n + \Delta_n)(x_0 + E)^n + (a_{n-1} + \Delta_{n-1})(x_0 + E)^{n-1} + \cdots \\ + (a_1 + \Delta_1)(x_0 + E) + (a_0 + \Delta_0) = 0.\end{aligned} \tag{4a}$$

If E is small in comparison with x_0, the first members of the development suffice, so that

$$\begin{aligned}(a_n + \Delta_n)(x_0^n + nEx_0^{n-1}) + (a_{n-1} + \Delta_{n-1})(x_0^{n-1} + (n - 1)Ex_0^{n-2}) + \cdots \\ + (a_1 + \Delta_1)(x_0 + E) + (a_0 + \Delta_0) = 0.\end{aligned} \tag{4b}$$

If this multiplication is carried out, we observe that x_0 satisfies the equation (4), and if the products are neglected, we get

$$\begin{aligned}E(na_nx_0^{n-1} + (n - 1)a_{n-1}x_0^{n-2} + \cdots + 2a_2x_0 + a_1) \\ + \Delta_nx_0^n + \Delta_{n-1}x_0^{n-1} + \cdots + \Delta_1x_0 + \Delta_0 = 0,\end{aligned} \tag{4c}$$

or

$$(5)\qquad |E| \leqq \frac{|\Delta_n x_0^n| + |\Delta_{n-1}x_0^{n-1}| + \cdots + |\Delta_1 x_0| + |\Delta_0|}{|f'(x_0)|}.$$

If now the absolute errors of the coefficients are approximately equal—as for example is the case if the errors arise from rounding off, and all the coefficients have the same number of decimal places—and if the largest of these is denoted by Δ, then

$$(6)\qquad \begin{aligned} |E|\cdot|f'(x_0)| \leqq |\Delta|\cdot\big|\,|x_0^n| + |x_0^{n-1}| + \cdots \\ + |x_0| + 1\,\big| = |\Delta|\left|\frac{|x_0^{n+1}| - 1}{|x_0| - 1}\right|. \end{aligned}$$

If, on the other hand, the coefficients are determined by observation, then, in general, the relative accuracy will be the same, and the relative error $\delta_m = \Delta_m/a_m$ must be introduced. In this case we have

$$(6a)\qquad x\cdot\epsilon\cdot f'(x_0) = a_n\delta_n x^n + a_{n-1}\delta_{n-1}x^{n-1} + \cdots + a_1\delta_1 x + a_0\delta_0\,.$$

If the maximum relative error is δ, it follows that

$$(7)\qquad |\epsilon| \leqq |\delta| \frac{|a_n x^n| + |a_{n-1}x^{n-1}| + \cdots + |a_1 x| + |a_0|}{|x\cdot f'(x)|}.$$

For very large values of x, this becomes $\delta a_n/na_{n-1}$.

4. *Example*: Let the coefficients of the equation

$$443x^2 - 226x - 515 = 0$$

be inaccurate to one half unit in the last place, because of rounding off. The inaccuracy of the two roots $x_1 = 1.363$, $x_2 = -0.853$ consequently amounts to

$$|E_1| \leqq \frac{|\Delta|}{|f'(x_1)|}\cdot\left|\frac{|x_1^3| - 1}{|x_1| - 1}\right| \approx \frac{0.5 \times 4.25}{982} \approx 0.0022,$$

$$|E_2| \leqq |\Delta|\left|\frac{|x_2^3| - 1}{|x_2| - 1}\right|\frac{1}{|f'(x_2)|} \approx \frac{0.5 \times 2.58}{982} \approx 0.0013.$$

Therefore

$$x_1 = 1.363 \pm 0.0022, \qquad x_2 = -0.853 \pm .0013.$$

If, in the same equation, the coefficients were known, by observation, to within about 1%, then the relative error of the first value of the root would be

$$|\epsilon_1| \leqq 0.01\frac{1648}{982 \times 1.363} \approx 0.0123 \approx 1\ 1/4\%,$$

and for the second value we would have

$$|\epsilon_2| \leqq 0.01 \frac{1030}{0.853 \times 982} \approx 0.0123 \approx 1\ 1/4\%.$$

If the error of the constant term in the example of Sec. 2 is assumed to be 0.05, then, since the other terms are correct,

$$E < \frac{0.05}{f'(x_0)} \approx \frac{0.05}{297} \approx 0.0002.$$

Therefore, the last digit can be in error by two units.

5. *Approximate values for the roots of an algebraic equation can be found graphically by Lill's method.* We first seek to draw two polygons in the framework of the equation. Between the endpoints of these, B_1 and $\bar{B}_1$, which correspond in Fig. 78 to the values $x = 0.95$ and $x = 1$, lies the point A_0. We then try, by interpolation, to draw the broken line between these two polygons whose endpoint B coincides with A_0. This then determines a root x_0 of the equation. This construction is materially simplified by the use of transparent millimeter paper, since then the polygons do not have to be drawn. Instead, they can be followed on the millimeter paper, appropriately rotated about the origin.

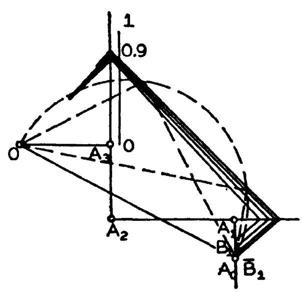


FIG. 78

According to the arrangements in 7.7, we can now use the polygon determining the root as the framework of the equation, from which the missing roots of the given equation can be determined. Of course the scale of measurement is λ times as large. After each root is found, the degree of the equation decreases by 1. Finally, an equation of second degree is obtained, where the structure consists of only three lines. A polygon of two lines is drawn between the beginning and end points. This construction can be carried out exactly with the circle of Thales, using the line connecting the two endpoints as diameter.

Example: In the figure, the equation

$$18.5x^3 + 16.1x^2 - 24.4x - 7.9 = 0$$

is represented. Two polygons, corresponding to $x = 0.95$ and $x = 1.0$, straddle the point A_0 with their endpoints. The interpolation gives $x = 0.98$, If the circle of Thales is drawn on OA_0 as a diameter, this gives the two dotted lines, from which $x_2 = -1.6$, $x_3 = -0.3$ are found.

6. The roots of algebraic equations can often be found with comparative ease, graphically as well as numerically, by use of *addition and subtraction logarithms*[3] (cf. 4.7). For the graphical solution, we first construct the "addition curve," i.e., to $\xi = \log t$ we plot $\eta = \log (1 + 1/t)$ as ordinate If we set $t = x_1/x_2$, then

$$\log (x_1 + x_2) = \log x_1 + \eta. \tag{8}$$

If therefore we have the lengths $\log x_1$ and $\log x_2$, we plot their difference as abscissa, and plot the corresponding ordinate on the greater length $\log x_1$; this gives the length of $\log (x_1 + x_2)$.

If x_1 is large in comparison to x_2, the curve ordinate becomes very small; for $x_1/x_2 \approx 1000$, this ordinate can safely be neglected in the drawing. But the relation also holds for $x_1 < x_2$, only then $\xi_1 = \log t_1$ is negative, and is therefore marked off from the origin to the left. To construct $\log (x_1 + x_2)$ graphically, the ordinate belonging to $\xi_1 = \log t_1$ is plotted on the length $\log x_1$. For the drawing of this portion of the addition curve, we observe that, for negative $\xi_1 = -\xi$,

$$\xi_1 = \log t_1 = -\log t = \log \frac{1}{t}, \qquad \text{i.e., } t_1 = \frac{1}{t},$$

$$\eta_1 = \log\left(1 + \frac{1}{t_1}\right) = \log (1 + t) = \log t\left(1 + \frac{1}{t}\right) \tag{9}$$

$$= \log t + \log\left(1 + \frac{1}{t}\right) = \xi + \log\left(1 + \frac{1}{t}\right) = \xi + \eta.$$

The positive x axis and the bisector of the second quadrant are therefore the asymptotes of the addition curve.

The construction of $\log (x_1 - x_2)$ from $\log x_1$ and $\log x_2$ can also be carried out with this curve, if we interchange the ordinate and abscissa (cf. 4.7), but it is better to draw a special *subtraction curve,* in which we mark off $\eta = \log (1 - 1/t)$ as ordinate to $\xi = \log t$. These ordinates are all negative. Their length, taken in absolute value, is then to be subtracted from $\log x_1$ to get a distance of length $\log (x_1 - x_2)$. This subtraction curve is symmetric with respect to the bisector of the fourth quadrant. If then we have two points of the curve (ξ_1, η_1) and (ξ_2, η_2), then by symmetry $\xi_2 = -\eta_1$, $\eta_2 = -\xi_1$ must be on this line. Therefore

$$\eta_2 = \log\left(1 - \frac{1}{t_2}\right) = -\xi_1 = -\log t_1 = \log \frac{1}{t_1}, \tag{10}$$

i.e.,

$$1 - \frac{1}{t_2} = \frac{1}{t_1}; \qquad t_2 = \frac{1}{1 - 1/t_1}, \tag{11}$$

from which it follows that

(12) $$\xi_2 = \log t_2 = -\log\left(1 - \frac{1}{t_1}\right) = -\eta_1 ,$$

so that the symmetry is demonstrated. Also, the negative η- and the positive ξ-axes are asymptotes. Fig. 79 shows the drawing of both curves,

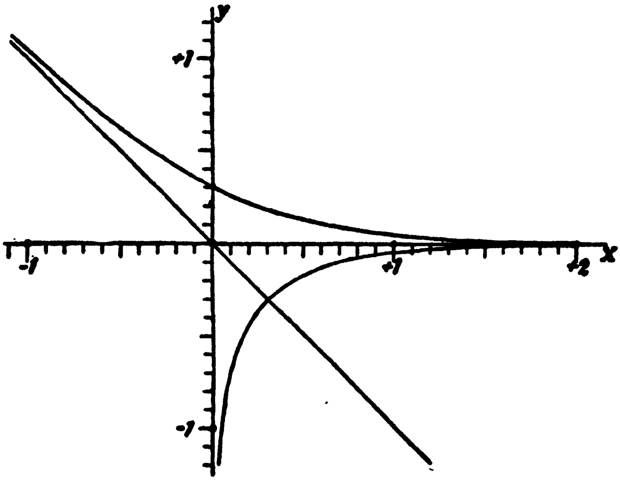


Fig. 79

with the help of which we now construct the logarithmic drawings of the function.

7. To construct the logarithmic graph of a rational integral function, we plot $\xi = \log x$ as abscissa and $\eta = \log y$ as ordinate. The graph of $y = a_m x^m$ is then a straight line:

$$\eta = m\xi + \log | a_m | .$$

Of course this representation is only possible for positive values of x. To draw a function

$$y = a_m x^m + a_n x^n$$

we plot the two straight lines

$$\eta = m\xi + \log | a_m |;$$

$$\eta = n\xi + \log | a_n |,$$

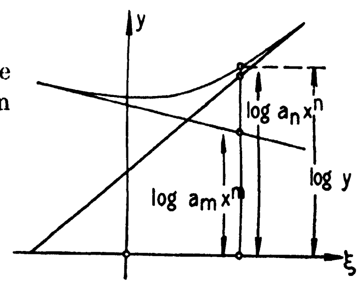


Fig. 80

nd draw the logarithmic curve from them by means of the addition or ıbtraction curves, according to whether a_m or a_n have the same or oppo-

site signs. The two straight lines are asymptotic to this curve. Here also we obtain curves only for $x > 0$, since the logarithm of a negative number is complex. If we also want the graph for $x < 0$, $-x$ must be substituted in the equation for x, and the curve of this equation constructed in the same way, whereby, under the circumstances, η is replaced by $-\eta$. Brauer has constructed a special compass with three points for carrying out the addition. If two points are set on the corresponding points of the two curves, then the third gives the desired point of the curve.[3]

If we have an expression of the form

$$y = a_m x^m + a_n x^n + a_p x_p$$

then we first plot the curve $\eta = \log (a_m x^m + a_n x^n)$ as above, and from this and the straight line $\eta = p\xi + \log |a_p|$, we plot the desired curve by use of the addition or subtraction logarithms. The graph of such an expression can be represented step-wise. Such an expression consists of a great number of these summands.

If the roots of such a function are to be determined, it is advisable to put the negative terms on the other side, so that the equation has the form

$$\varphi(x) = \psi(x)$$

where both functions, which now contain only positive terms, are to be constructed by the above method. Mehmke, who first worked with logarithmic graphs, has shown that one obtains two curves concave upwards

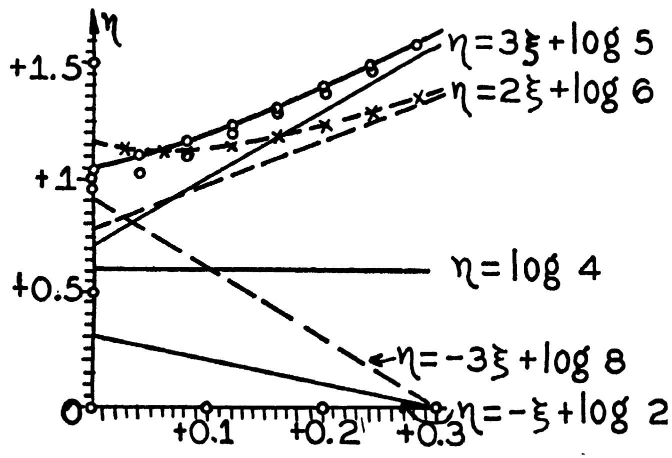


FIG. 81

for positive η, the intersection of which gives the desired approximate value of the roots.[4] Now and then, better intersections are obtained with another distribution of the summands, but there is no general rule.

8. *Example*: The equation

$$y = 5x^6 - 6x^5 + 4x^3 + 2x^2 - 8 = 0$$

has a root between $x = 1$ and $x = 2$. The equation is broken up and divided by x^3 to avoid too steep a curve, so that

$$5x^3 + 4 + \frac{2}{x} = 6x^2 + \frac{8}{x^3}.$$

The curves are drawn in Fig. 81 between $\xi = 0$ and $\xi = 0.301$. For the left side, $5x^3 + 4$ is plotted first; this gives the curve marked by the unconnected circles. From this we get the drawn curve by addition of $2/x$. The curve representing the right side is dotted. As an approximate value we get

$$\xi = 0.056, \qquad \text{i.e., } x = 1.137.$$

For numerical calculation *by use of the addition logarithms*, Runge[5] has rearranged the *approximation methods* given in this section.

9. Finally, we note that apparatus has been constructed for the mechanical determination of the roots of an algebraic equation—the so-called equation machine or carriage. This is a mechanism on which forces are acting, the magnitudes of which are determined by the coefficients of the equation. The equilibrium position is then determined by the action of these forces. The roots of the equation can be read off from such a position.

Ordinary carriages, systems of such devices, rotation machines, hydrostatic machines, electrical machines, which are based on Kirchhoff's law, or electromagnetic devices, which use the magnetic fields of electric currents for the determination of complex roots, are all used. All these devices provide very rough approximation values, which can then be improved by one of the methods of the preceding sections. Further information on these mechanisms can be found in the literature.[6]

NOTES

1. Ruffini, *Sopra la determinazione delle radici* (Modena, 1804). Horner, *Phil. Trans.* (1819), I, p. 308.

2. Mehmke, *Leitfaden zum graphischen Rechnen* (Leipzig, 1917); *Zivilingenieur*, **35** (1889).

3. Dyck, *Katalog math. u. math.-phys. Modelle, Apparate u. Instrumente.* Appendix (Munich, 1893), p. 40; Pfeiffer, *Z. f. angew. Math. u. Mech.* **5** (1925), p. 172.

4. Mehmke, *Leitfaden zum graphischen Rechnen* (Leipzig, 1917).

5. Runge, *Praxis der Gleichungen* (Leipzig, 1900), p. 127.

6. Mehmke, *Enzykl. d. math. Wissenschaft* I, 2. Numerisches Rechnen; Jacob, *Calcul mécanique* (Paris, 1911); Riebesell, *Z. f. Math. u. Phys.* **63** (1915), p. 256.

20. The Number and Position of the Real Roots of an Equation.

1. Approximate values of the roots of the equation must be known beforehand for all the methods described in the two preceding articles. In order to find these, we enclose them between definite limits, i.e., the number of real roots lying in a given interval is determined. As a criterion for this, we use the number of the *sign sequences and sign changes* the particular function series has at the ends of the observed interval. If we have a sequence of functions

$$f(x), \qquad g(x), \qquad h(x) \cdots, \tag{1}$$

and if we substitute for x a particular value a in all such functions, and consider only the signs of the individual function values, then we have a sign sequence if two successive function values have the same sign. We have a sign reversal if they have opposite signs.

2. By use of this concept, a rule is put forward by Budan and Fourier which gives an upper limit for the number of the roots lying in an interval. By this rule, the sequence

$$f(x), \qquad f'(x), \qquad f''(x) \cdots f^{(n)}(x) \tag{2}$$

is to be formed, if we seek the number of the real roots of an equation

$$f(x) = 0 \tag{2a}$$

in an interval from $x = a$ to $x = b$, where a and b are not roots of $f(x)$. We then substitute $x = a$, $x = b$ in this sequence. In this, $f^{(n)}(x)$ must not change its sign in the observed interval. In a rational integral function of nth degree, this is obviously so, since then $f^{(n)}(x) = c$ is a constant. But the theorem also holds for an analytic function if only $f^{(n)}(x) = c$ does not change sign. The rule may be stated as follows. If the above conditions are fulfilled, then

(a) *the number of sign reversals* in the above sequence *can only decrease with increasing* x,

(b) *the number of sign reversals lost in an interval is equal to the number of the real roots lying in this interval or exceeds it by an even number.*

To derive this rule, we set forth the following considerations:

(a) If x_0 is a simple root of $f(x) = 0$, then for sufficiently small h, if the curve intercepts the X axis increasing or decreasing,

(3)

	f	f'
$x_0 - h$	$-$	$+$
$x_0 + h$	$+$	$+$

	f	f'
$x_0 - h$	$+$	$-$
$x_0 + h$	$-$	$-$.

Therefore, in the beginning of the above function sequence, a sign reversal is always lost whenever x passes through a real root of the function $f(x)$. With a double root, two sign reversals are lost, with a triple, three, etc. This can be considered exactly as above, but it is also a consequence of the following considerations.

(b) If x_0 is a root of m consecutive functions

$$f^{(r-m+1)}(x), \qquad f^{(r-m+2)}(x), \cdots, f^{(r)}(x), \tag{4}$$

then, if we expand in a Taylor series,

$$f^{r}(x_0 + h) = hf^{(r+1)}(x_0) + \frac{h^2}{2!}f^{(r+2)}(x_0) + \frac{h^3}{3!}f^{(r+3)}(x_0)\cdots,$$

$$f^{(r-1)}(x_0 + h) = \frac{h^2}{2!}f^{(r+1)}(x_0) + \frac{h^3}{3!}f^{(r+2)}(x_0) + \cdots, \tag{5}$$

$$f^{(r-2)}(x_0 + h) = \frac{h^3}{3!}f^{(r+1)}(x_0) + \frac{h^4}{4!}f^{(r+2)}(x_0) + \cdots$$

. . . .

If h is sufficiently small, then the first term determines the sign of the right side. Therefore, for positive h, all functions $f^{(r)}(x)$ to $f^{(r-m+1)}(x)$ have the same sign in the immediate neighborhood of the root, while they have alternating signs for small negative h. All all events, sign reversals are lost with increasing x. For more exact investigation, we distinguish two cases:

(α) *Let m be an even number.* Let us take as an example, $m = 4$; for other values of m, the result follows directly. Further, let us assume that $f^{(r+1)}(x_0)$ is positive. This is not a serious restriction, because if $f^{(r+1)}(x_0)$ is negative, we simply give the entire function sequence the opposite sign. We have then the following signs:

(6)

	$f^{(r-4)}$	$f^{(r-3)}$	$f^{(r-2)}$	$f^{(r-1)}$	$f^{(r)}$	$f^{(r+1)}$
$x_0 - h$	$\pm$	$+$	$-$	$+$	$-$	$+$
x_0	$\pm$	0	0	0	0	$+$
$x_0 + h$	$\pm$	$+$	$+$	$+$	$+$	$+$,

since $f^{(r-m)}(x_0)$ can be positive as well as negative. There are four sign reversals lost. In general, with even m, exactly m sign reversals are lost.

(β) *Let m be odd,* say $m = 3$. Then we have, under the previous assumptions,

(7)

	$f^{(r-3)}$	$f^{(r-2)}$	$f^{(r-1)}$	$f^{(r)}$	$f^{(r+1)}$
$x_0 - h$	$\pm$	$-$	$+$	$-$	$+$
x_0	$\pm$	0	0	0	$+$
$x_0 + h$	$\pm$	$+$	$+$	$+$	$+$.

According as $f^{(r-3)}$ and $f^{(r+1)}$ have the same or opposite signs, we lose here four or two sign reversals. In the first case, the one sign reversal shifts one place to the left. Therefore the following is obtained:

If x_0 is a simple root of $f(x) = 0$, a sign reversal is lost at this place in the above function sequence.

If x_0 is a root of m derived functions, then with *even m, m, with odd m either $m - 1$ or $m + 1$ are lost. Hence an even number of sign reversals is always lost.*

In particular, if for each value of x for which a derivative $f^{(r)}(x)$ is zero, the preceding $f^{(r-1)}(x)$ and the following $f^{(r+1)}(x)$ are not zero, and have opposite signs, no sign reversal is lost in the sequence. Only if $r = 0$, if therefore $f(x) = 0$, is a sign reversal lost. The number of roots in this case is equal to the number of lost sign reversals.

3. For an algebraic equation, the above function sequence has in general no sign reversal for $x = \infty$, since all derivatives have the sign of the term of highest power in x. For $x = -\infty$, there are only sign reversals, and no sign sequences. If, therefore, no sign reversal is lost in the interior of the above function sequence, then n sign reversals must have been lost at the beginning of the interval. But this is only possible if the equation has n real roots.

For $x = 0$, the number of sign sequences and sign reversals of the series of derivatives coincides with that of the series of coefficients. From this is derived the *rule of Descartes or Harriot*: *The number of negative real roots of an algebraic equation is equal to the number of sign sequences among the coefficients, or is smaller by an even number; the number of positive real roots is equal to the number of sign reversals of this sequence or is smaller by an even number.*

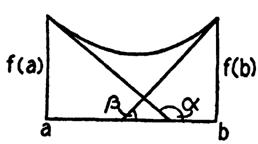


Fig. 82

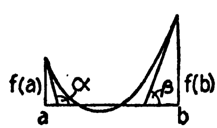


Fig. 83

4. The following proceeds from Fig. 82 and Fig. 83. If we know that no inflection point of the curve lies between a and b, that therefore $f''(x)$

is not zero, and if we have chosen the interval sufficiently small, then with two sign reversals lost,

A. *no real roots* lie in the interval if

$$|\, b - a \,| < -\frac{f(a)}{\operatorname{tg}\alpha} + \frac{f(b)}{\operatorname{tg}\beta} = -\frac{f(a)}{f'(a)} + \frac{f(b)}{f'(b)}; \tag{8}$$

B. *possibly two real roots* lie in the interval if

$$|\, b - a \,| > -\frac{f(a)}{f'(a)} + \frac{f(b)}{f'(b)}. \tag{9}$$

5. *Example*: By Descartes' rule, the equation

$$f(x) = x^6 - 5x^5 + 3x^4 + 10x^3 - 8x^2 - 4x + 1 = 0 \tag{9a}$$

can have two real negative roots because of two sign sequences and four real positive roots because of four sign reversals.

To obtain further details about the position and number of the positive roots, we develop $f(x)$ in a series in powers of $u = (x - 1)$, $v = (u - 1) = (x - 2)$, $w = (v - 1) = (x - 3)$, etc., according to Horner's scheme. The number of sign reversals lost in each succeeding development permits a conclusion as to the number of roots lying between $x = 0$ and $x = 1$, or $x = 1$ and $x = 2$, etc. Since we must always multiply by the factor 1, the number in the preceding column and one row lower must always be added to each coefficient. If we form this scheme, it appears as follows, for example, for $x = 1$,

(9b)

1	−5	+3	+10	−8	−4	+1
1	−4	−1	+ 9	+1	−3	$-2 = f(1)$
1	−3	−4	+ 5	+6	$+3 = f'(1)$	
1	−2	−6	− 1	$+5 = f''(1)/2!$		
1	−1	−7	$-8 = f'''(1)/3!$			
1	0	$-7 = f^{(4)}(1)/4!$				
1	$1 = f^{(5)}(1)/5!$					
$1 = f^{(6)}(1)/6!$						

For $x = 4$, we have only sign sequences in x. Therefore the equation can have no roots which are larger than 4. We have the sequences of signs as follows:

$x = 0$	$+ - - + + - +$	(4 changes)	1 real root
$x = 1$	$- + + - - + +$	(3 changes)	
$x = 2$	$- - - - + + +$	(1 change)	possibly two real roots

$$\begin{array}{llll} x = 3 & - - + + + + + & \text{(1 change)} & \\ & & & \text{1 real root} \\ x = 4 & + + + + + + + & \text{(no changes).} & \end{array}$$

For the investigation of the negative roots, we replace x by $-x$ and consider the equation

$$x^6 + 5x^5 + 3x^4 - 10x^3 - 8x^2 + 4x + 1$$

in exactly the same way. This gives

$x = 1$:	1	5	3	-10	-8	$+4$	$+1$
	1	6	9	-1	-9	-5	$-4 = f(-1)$
	1	7	16	$+15$	$+6$	$-1 = -f'(-1)$	
	1	8	24	39	$+45 = f''(-1)/2!$		
	1	9	33	$72 = f'''(-1)/3!$			
	1	10	$43 = f^{(4)}(-1)/4!$				
	1	$11 = f^{(5)}(-1)/5!$					
	$1 = f^{(6)}(-1)/6!$						
$x = 2$:	1	11	43	72	45	1	-4
	1	12	55	127	172	173	$169 = f(-2)$.

The calculation thus far suffices, because we see that no negatives values can appear, that therefore in this series of terms, only sign sequences occur. Therefore we have for $f(x)$

$$\begin{array}{llll} x = 0 & + - - + + - + & \text{(4 changes)} & \text{1 real root} \\ x = -1 & + - + - + - - & \text{(5 changes)} & \\ x = -2 & + - + - + - + & \text{(6 changes).} & \text{1 real root} \end{array}$$

There still remains to be investigated whether real roots lie between 1 and 2. Since $f''(x)$ changes its sign in the interval, the criterion (8) cannot be used. We must therefore consider a smaller interval. We begin from $x = 1$ and expand about 1.5:

$$f(1.5) = -0.640625, \quad f'(1.5) = -1, \quad f''(1.5) = -15.3125 \times 2!,$$

$$f'''(1.5) = -17 \times 3!, \; f^{(4)}(1.5) = -0.75 \times 4!$$

$$f^{(5)}(1.5) = +4 \times 5!, \; f^{(6)}(1.5) = 1 \times 6!$$

Since only one sign reversal is lost in the interval, the roots can only lie between 1 and 1.5. Here also we cannot use the above criterion, since $f''(x)$ also changes its sign in this interval, and therefore becomes zero somewhere in the interval. We therefore make the development about 1.4 by means of the Horner scheme:

$$f(1.4) = -0.676864,\ f'(1.4) = +1.55744,\ f''(1.4) = -10.296 \times 2!,$$
$$f'''(1.4) = -16.32 \times 3!,\ f^{(4)}(1.4) = -2.6 \times 4!$$
$$f^{(5)}(1.4) = 3.4 \times 5!,\ f^{(6)}(1.4) = 1 \times 6!$$

Here we still have the three sign reversals, and the second derivative does not change its sign in the interval. The roots must therefore lie between 1.4 and 1.5 if they are real. But if we use the criterion above,

$$\frac{f(1.5)}{f'(1.5)} - \frac{f(1.4)}{f'(1.4)} = 0.641 + 0.435 > 1.$$

Therefore there are no real roots in this interval.

There is *another series of theorems which gives an upper limit for the number of real roots in an interval*, but this contributes nothing essentially different from the above.[1]

6. More useful than all these theorems is the *Sturm series*,[2] which gives the exact number of roots in an interval. The functions

$$(10) \qquad f(x), \quad f_1(x), \quad f_2(x), \quad \cdots, \quad f_n(x)$$

of such a series must have the following properties:

I. All functions must have only simple roots, and must not vanish at the limits of the interval;

II. The second function of the sequence, $f_1(x)$, must have the same sign as $f'(x)$ at all roots of $f(x)$;

III. The last function of the sequence, $f_n(x)$, must not change its sign in the interval;

IV. At points where one of the functions vanishes, the preceding and following functions must have opposite signs.

Such a sequence can be formed in the following way. First we find the derivative $f'(x)$ of $f(x)$. Then $f(x)$ is divided by $f'(x)$. This gives the quotient $g_1(x)$ and the remainder $-r_1(x)$. Then

$$(11) \qquad f(x) = f'(x) \cdot g_1(x) - r_1(x).$$

If we divide $f'(x)$ by $r_1(x)$, we get the quotient $g_2(x)$ and the remainder $-r_2(x)$, etc. We then proceed as in the investigation of the least common divisor, or in the development of non-integral rational functions in continued fractions, except that the remainder is written with a negative sign. We then have

$$f'(x) = r_1(x) \cdot g_2(x) - r_2(x),$$
$$(12) \qquad r_1(x) = r_2(x) \cdot g_3(x) - r_3(x),$$
$$\cdots\cdots\cdots\cdots$$

In the following, we consider only rational integral functions. With these, the degree of the function decreases by one in each successive term. Then $g_m(x)$ will be a linear function.

7. There are now two possibilities. If $f(x)$ and $f'(x)$ have a common divisor $\varphi(x)$, if therefore, $f(x)$ has multiple roots, then $r_1(x)$ must also have this divisor. Then $r_2(x)$ must also have the factor $\varphi(x)$. This process therefore must lead to this divisor with rational integral functions. By division with $\varphi(x)$, *$f(x)$ and $f'(x)$ can then be made prime to each other.* Then $f(x)$ will no longer have multiple roots. In this case, $r_n(x)$ will be a constant. If it can easily be determined that an earlier remainder $r_m(x)$ does not change its sign in the interval considered, then the sequence can be broken off with this term. Further division is then unnecessary.

The series

(13) $$f(x), \quad f'(x), \quad r_1(x), \quad \cdots, \quad r_n(x)$$

is then a Sturm series. It can be seen immediately that by suitable choice of the boundaries of the interval, the condition I-III can be satisfied. But then IV is also. Since $f(x)$ has only simple roots, by II, a sign reversal is always lost if a root is passed through with increasing x. If, on the other hand, x_0 is a root of $r_p(x)$, then it cannot also be a root of the adjacent functions $r_{p-1}(x)$ and $r_{p+1}(x)$; because, if, for example, it were a root of $r_{p+1}(x)$, then, since

(14) $$r_{p-1}(x) = g_{p+1}(x)r_p(x) - r_{p+1}(x_0),$$

it must also be a root of $r_{p-1}(x)$, and consequently also of $r_{p-2}(x)$, $r_{p-3}(x)$, $\cdots$, $r_1(x)$, $f'(x)$, $f(x)$; i.e., $f(x)$ must have a multiple root at this point, and this is not the case. For $r_p(x_0) = 0$, we have

(15) $$r_{p-1}(x_0) = -r_{p+1}(x_0).$$

The adjacent functions have opposite signs. If h is now chosen so small that the functions $r_{p-1}(x)$ and $r_{p+1}(x)$ do not change their signs between $x_0 - h$ and $x_0 + h$, then we have the following two possibilities already contained in Sec. 2(b):

(16)

	r_{p-1}	r_p	r_{p+1}	r_{p-1}	r_p	r_{p+1}
$x_0 - h$	+	$\pm$	$-$	$-$	$\pm$	+
x_0	+	0	$-$	$-$	0	+
$x_0 + h$	+	$\mp$	$-$	$-$	$\mp$	+

Therefore, the sequence never loses a sign reversal at such a root, as was shown in 20.2. The sign reversal is merely displaced. Now, since $r_n(x)$ does not change its sign, then the sequence can only lose a sign reversal when the value of x passes through a root.

8. *Example*: The position of the real roots of the equation

(16a) $f(x) = 5x^5 - 14.32x^4 + 4.53x^3 - 15.39x^2 - 9.86x + 23.97 = 0$

is to be determined. From the equation we obtain

(16b) $f'(x) = 25x^4 - 57.28x^3 + 13.59x^2 - 30.78x - 9.86,$

and by division,

$$\begin{aligned} r_1(x) &= 4.750x^3 + 7.677x^2 + 11.414x - 22.84, \\ r_2(x) &= -111.40x^2 - 324.17x + 479.58, \\ r_3(x) &= -49.75x + 49.30, \\ r_4(x) &= -48.95. \end{aligned}$$

This gives the following sign table:

	$f(x)$	$f'(x)$	$r_1(x)$	$r_2(x)$	$r_3(x)$	$r_4(x)$	
$-\infty$	$-$	$+$	$-$	$-$	$+$	$-$	
-1	$-$	$+$	$-$	$+$	$+$	$-$	1 Root
0	$+$	$-$	$-$	$+$	$+$	$-$	1 Root
$+1$	$-$	$-$	$+$	$+$	$-$	$-$	
$+2$	$-$	$-$	$+$	$-$	$-$	$-$	1 Root
$+3$	$+$	$+$	$+$	$-$	$-$	$-$	
$+\infty$	$+$	$+$	$+$	$-$	$-$	$-$	

The equation then has 3 real roots.

The spherical harmonics also form a Sturm series, for which the equation

(17) $(n + 1)P_{n+1}(x) - (2n + 1)P_n(x) + nP_{n-1}(x) = 0$

holds. This permits conclusions as to the real nature of the roots between $+1$ and -1 (16.11; 27.8; 27.9).[3]

9. If the *difference scheme* of an algebraic function has been formed, then, without further construction, conclusions can be drawn as to the position of the roots of this function.

From Newton's formula, 10(11),

$$y = y_0 + \frac{t}{1!}\Delta^1_{1/2} + \frac{t(t-1)}{2!}\Delta^2_1 + \frac{t(t-1)(t-2)}{3!}\Delta^3_{3/2} \cdots \tag{18}$$
$$\frac{t(t-1)\cdots(t-n+1)}{n!}\Delta^n_{n/2},$$

it can be seen that if y_0 and all the Δ's have the same sign for values of $t > n - 1$, y cannot change its sign, since all polynomials with which

the differences are multiplied are then positive. In this case therefore, the equation $y = f(x) = 0$ can have no roots which are larger than $x_0 = (n - 1)h$.

If, on the other hand, y_0, $\Delta^1_{1/2}$, Δ^2_1, $\cdots$, $\Delta^n_{n/2}$ have alternating signs, then y cannot change its sign for $t < 0$, since then the polynomials also must have alternating signs, and therefore all summands of the formula N_+ must have the same sign. Then the equation $f(x)$ can have no roots which are smaller than x_0.

If we consider the formula 10(13).

$$N_-(t) = y_0 + \Delta^1_{-(1/2)}t + \Delta^2_{-1}\frac{t(t+1)}{2!} + \cdots + \Delta^n_{-(n/2)}\frac{t(t+1)\cdots(t+n-1)}{n!}, \tag{19}$$

we find that for like signs on y_0 and the differences, no root of the equation can be larger than x_0, while with alternating signs, no root can be smaller than $x_0 - (n - 1)h$.

If we write the formula (18) in the form

$$N_+(t) = y_0 + \frac{t}{1}\Delta^1_{1/2} - \frac{t(1-t)}{2!}\Delta^2_1 + \frac{t(1-t)(2-t)}{3!}\Delta^3_{3/2}\cdots (-1)^{n-1}\frac{t(1-t)(2-t)\cdots(n-1-t)}{n!}\Delta^n_{n/2}, \tag{20}$$

then in the interval $x_0 \leqq x < x_0 + h$, i.e., $0 \leqq t \leqq 1$, all polynomial coefficients of the differences are positive, and it is certain that

$$\frac{t(1-t)(2-t)\cdots(m-t)}{(m+1)!} < \frac{1}{m+1}. \tag{20a}$$

If therefore, the differences $\Delta^a_{a/2}$, $\Delta^b_{b/2}$, $\cdots$ have opposite signs to y_0, then $f(x)$ cannot be zero in the interval, provided that

$$|y_0| > \left|\frac{\Delta^a_{a/2}}{a} + \frac{\Delta^b_{b/2}}{b}\cdots\right|. \tag{20b}$$

NOTES

1. Fricke, *Algebra I*, Sec. II (Braunschweig, 1924).
2. Sturm, *Bulletin de Férussac*, **11** (1829).
3. Jacobi, *Crelle's Journal* **II** (1827), p. 223.

21. Position and Approximate Determination of the Complex Roots of an Algebraic Equation.

1. If, in the rational integral function

(1) $$g(z) = c_n z^n + c_{n-1} z^{n-1} + \cdots + c_1 z + c_0 ,$$

the coefficients of which may be in general complex numbers, we introduce polar coordinates $z = re^{i\varphi}$ for z, the equation becomes

(2) $$g(z) = c_n r^n e^{in\varphi} + c_{n-1} r^{n-1} e^{i(n-1)\varphi} + \cdots + c_1 re^{i\varphi} + c_0 ,$$

or, if we also set

(3) $$g(z) = C_n r^n e^{i(n\varphi + \alpha_n)} + C_{n-1} r^{n-1} e^{i[(n-1)\varphi + \alpha_{n-1}]} + \cdots + C_1 re^{i(\varphi + \alpha_1)} + C_0 e^{i\varphi_0}.$$

The complex number is represented by a particular point Q_m of the Gaussian complex plane, or by the vector OQ_m from the origin to this point. The factor $e^{im\varphi}$ represents a rotation of the vector through the angle $m\varphi$. This then gives the vector OP_m and the sum of all these vectors is

(4) $$g(z) = \sum_{m=0}^{m=n} OP_m = OP = Re^{i\Phi} = X + iY.$$

If z_0 is a root of this equation, then $g(z_0) = 0$; i.e., $X(x_0, y_0) = 0$ as well as $Y(x_0, y_0) = 0$. Since $\operatorname{tg} \Phi = X/Y$, Φ is indeterminate at such a point.

To investigate *the path of the curve* $\Phi = c$ *in the neighborhood of such a root* z_0, we develop $g(z)$ in powers of $z - z_0$:

(4a) $$g(z) = g(z_0) + g'(z_0)(z - z_0) + \frac{1}{2!} g''(z_0)(z - z_0)^2 + \cdots + \frac{1}{n!} g^{(n)}(z_0)(z - z_0)^n.$$

Now if z_0 is a p-fold root of $g(z)$, then $g(z_0)$ and the first $p - 1$ derivatives $g'(z_0) = \cdots = g^{(p-1)}(z_0) = 0$. If $z - z_0 = \rho e^{i\psi}$, $g^{(l)}(z_0) = D_l e^{i\beta_l}$, then

(5) $$g(z) = \sum_{l=p}^{l=n} \frac{D_l}{l!} \rho^l e^{i(l\psi + \beta_l)}.$$

If ρ is very small, i.e., if we consider a very small circle about this root, then the first term of this expansion predominates, and

(6) $$g(z) \approx \frac{D_p}{p!} \rho^p e^{i(p \cdot \psi + \beta_p)} = \overline{R} e^{i\overline{\Phi}},$$

or

(7) $$\overline{R} \approx \frac{D_p \cdot \rho^p}{p!}; \qquad \overline{\Phi} = p \cdot \psi + \beta_p .$$

Therefore, if ψ increases from 0 to 2π, $\bar{\Phi}$ increases by $2\pi p$. The *p-fold root is then a source point of the line $\Phi = c$ of $2\pi p$-fold strength.* There are no source points other than the roots of $g(z) = 0$, as may be seen if we set $p = 0$.

Therefore, if any region of the complex plane is bounded by a closed curve, the value of the integral $T = 1/2\pi \int d\Phi$ over this curve is equal to the number of complex roots lying in this region.

2. If we start with the development (3),

$$g(z) = C_n r^n e^{i(n\varphi + \alpha_n)} + C_{n-1} r^{n-1} e^{i[(n-1)\varphi + \alpha_{n-1}]} + \cdots + C_1 r e^{i(\varphi + \alpha_1)} + C_0 e^{i\alpha_0}, \tag{7a}$$

then the first term predominates for sufficiently large r, i.e., on a circle with a very large radius,

$$\Phi = n\varphi + \alpha_n . \tag{7b}$$

Since Φ increases by $2n\pi$ in one rotation, n and only n roots of the algebraic equation lie inside. There can be no roots outside, because no matter how large we choose r, Φ always increases by $2n\pi$ in traversing the circle (Gauss' theorem). It can easily be seen that no roots can lie outside of a circle with radius

$$R = (C_{n-1} + C_{n-2} + \cdots + C_0) : C_n ; \tag{8}$$

for

$$g(z) = z^{n-1}\left(c_n z + c_{n-1} + \frac{c_{n-2}}{z} + \cdots + \frac{c_0}{z^{n-1}}\right). \tag{9}$$

But for $|z| > 1$, by (8),

$$\left| c_{n-1} + \frac{c_{n-2}}{z} + \cdots + \frac{c_0}{z^{n-1}} \right| < R \cdot C_n . \tag{10}$$

Therefore, for $z > R$, the expression in parentheses in (9) cannot vanish, and no root of $g(z)$ can lie outside of this circle.

3. Since

$$X = R \cos \Phi; \qquad Y = R \sin \Phi, \tag{10a}$$

we can determine from the sign changes of X and Y how often the argument Φ increases by 2π. If we consider X and Y as functions of Φ, then we can regard $-Y$ as the derivative of X. If Φ increases, then we have the following signs in the respective quadrants:

Φ	1.	2.	3.	4.	5.	6.	$\cdots$ Quadrant
X	+	−	−	+	+	−	$\cdots$
$-Y$	−	−	+	+	−	−	$\cdots$

A Sturm series is then formed from X, $-Y$. The series loses a sign change whenever X passes through zero with increasing Φ, and gains a sign change whenever X passes through zero with decreasing Φ. If, in the circuit of a closed curve in the complex plane, X changes its sign l times in which the above series loses a sign change, and m times in which it gains a sign change, then the argument increases by $(l - m)\pi$, i.e., in the interior of a closed curve traversed in counter-clockwise fashion there are $(l - m)/2$ roots of the equation $g(z) = 0$, each root counted according to its multiplicity.

4. Since Φ is approximately $n\varphi + \alpha_n$ for circles of very large radius, then on half of such a circuit, Φ increases by $n\pi$. Such a circle is now divided by a straight line given in parametric form:

$$x = c + dt, \qquad y = c' + d't. \tag{10c}$$

We consider the region which is bounded by the straight line and the semi-circle—that region which is traversed in a counter-clockwise direction, i.e., the region which lies on the left side of the direction of passage of the straight line. If the Sturm series formed from X and $-Y$ loses l and gains m sign changes in passing along the straight line, then Φ must increase by $(l - m)\pi$. Therefore

$$\frac{n}{2} + \frac{l - m}{2}$$

roots must lie in the region on the left side of the line. Then $l - m$ *is the difference in the number of sign changes in the Sturm series for* $t = -\infty$ *and* $t = +\infty$, *or* $t = +\infty$ *and* $t = -\infty$, according as the number of roots on one side or the other of the straight line is sought.

In equations with real coefficients, the computation is simplified if it is a case of the number of roots on one side of a perpendicular to the real axis: $x = a$, $y = t$ or $z = a + it$. This case plays a role in technical problems. In vibration problems, for example, it is always an important question as to whether or not the oscillations in the system are damped, i.e., whether the real parts of all the roots of the algebraic equation determining the frequency of oscillation are negative. This can be determined by the above method.

5. *Example*: The equation

$$100z^4 + 63z^3 + 62z^2 + 7.3z + 0.6 = 0$$

is obtained for the transverse vibrations of the airplane considered by Deimler[4] (v. 18.4). In this case, the velocity is $r = 10\text{m/sec}$. The machine is stable if the real parts of the roots are negative, i.e., lie to the left of the imaginary axis $z = it$. If we substitute this, we obtain

$$100t^4 - 63it^3 - 62t^2 + 7.3it + 0.6 = 0$$

$$X = 100t^4 - 62t^2 + 0.6; \; -Y = 63t^3 - 7.3t.$$

The Sturm series is then formed:

$$r_1 = 50.41t^2 - 0.6, \qquad r_2 = 6.55t, \qquad r_3 = 0.6.$$

From this there results

	X	$-Y$	r_1	r_2	r_3
$-\infty$	+	−	+	−	−
$+\infty$	+	+	+	+	+

Therefore the number of sign changes lost is $l - m = 4$, so that the number of roots with negative real parts is

$$\frac{l-m}{2} + \frac{n}{2} = 2 + 2 = 4;$$

i.e., the equation has only roots with negative real parts. The vibrations of the airplane are therefore damped.

To determine how great the damping is, i.e., how large are the real parts of the roots, the investigation can be carried out for any suitably chosen parallel to the imaginary axis, e.g., for $z = -0.25 + it$. If we perform a Horner development for $t_0 = -0.25$ and substitute, then the Sturm series becomes

$$X = 100t^4 - 52.25t^2 + 2.056$$

$$-Y = -37t^3 + 18.14t$$

$$r_1 = 3.23t^2 - 2.056$$

$$r_2 = 5.43t$$

$$r_3 = 2.056.$$

Then

	X	$-Y$	r_1	r_2	r_3
$-\infty$	+	+	+	−	+
$+\infty$	+	−	+	+	+ .

Since no sign changes are gained or lost, $l - m$ is zero, i.e., only two roots lie on the negative side of this line, and the real parts of the others lie between -0.25 and 0. But the real part of the root on the left lies near -0.25. Then for

$$z = -0.26 + it$$

we get, just as above,

$$X = 100t^4 - 53.42t^2 + 2.243$$

$$-Y = -41t^3 + 19.19t$$

$$r_1 = 6.62t^2 - 2.243$$

$$r_2 = -5.29t$$

$$r_3 = 2.243.$$

From this follows

	X	$-Y$	r_1	r_2	r_3
$-\infty$	+	+	+	+	+
$+\infty$	+	−	+	−	+

Therefore $l - m = -4$. No roots lie to the left of this line, since

$$\frac{l-m}{2} + \frac{n}{2} = -2 + 2 = 0.$$

The real parts of the first pair of roots can be bounded in a similar way.

Hurwitz[5] has devised *another criterion* for determining whether an *equation with real coefficients* has only roots whose real parts are negative.

6. We saw in Sec. 1 of this article that the roots are the source points of the lines $\Phi = c$. If these lines are drawn at intervals $\Delta c = \pi/4$, then such lines emanate from an l-fold root. The drawing of these lines is materially simplified by an *equation machine* constructed by *Thommeck.*[6] This device permits the operator to read off the value Φ corresponding to

each pair of values r, φ, for $z = re^{i\varphi}$. The principle of the device is based on the following theorem:

"If $n + 1$ vectors OP_0, OP_1, $\cdots$, OP_n drawn from a point O have the resultant vector OP (equation 4), and if n identical mass points are placed at the points P_0, P_1, $\cdots$ P_n, whose center of mass is S, the three points O, S, P are collinear, and the ration OS/OP is always $1/(n + 1)$."

The theorem is obviously true for two vectors OP_0 and OP_1. We place at S' the mass 2 whose moment about each axis through O is equal to that of a mass 1 at P'. If the mass 1 is put at P_2, then the mass 2 at Q'' or 1 at P'' or 3 at S'' has the same moment about an axis through O if $OS'' = 2/3\ OQ'' = 1/3\ OP'$, as the masses 1 at P_2 and P', or at P_0, P_1 and P_2. The above theorem would then hold for 3 vectors. In general the theorem may then be proved by induction. If therefore we place masses of equal magnitude at the points P_0, P_1, $\cdots$, P_n, mentioned in 1, with the coordinates c_0, $c_1re^{i\varphi}$, $\cdots$, $c_nr^ne^{in\varphi}$ then the argument Φ of the center mass S is the same as that of the resultant vector $g(z) = Re^{i\Phi}$.

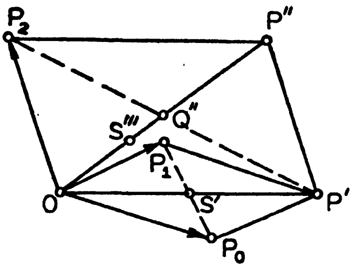


FIG. 84

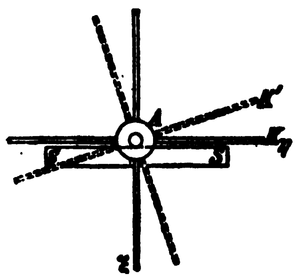


FIG. 85

7. The *Thommeck apparatus* consists of an axis A which can roll back and forth on two horizontal knife edges SS. On this axis, $n + 1$ hubs are mounted, each of which bears a cross K of 4 mutually perpendicular rods, which lie in a plane perpendicular to the axis. The first hub is rigidly connected to the axis, while each succeeding hub can be rotated through an angle φ with respect to the preceding. This is done by means of a cogwheel mechanism, so that the third forms an angle of 2φ with the first, the fourth an angle 3φ, and the $(n + 1)$st the angle $n\varphi$. This adjustment is made possible by means of a worm gear on the last hub. The others are then correctly placed by the cogwheel. The whole arrangement must be carefully assembled so that it is in neutral equilibrium at each position. This may be done by using small additional weights. Four equal movable weights are placed on the four spokes on each hub, one on each spoke. These weights are so placed on the spokes of the first hub that its center of mass has, in a suitably chosen scale, the coordinates

$$c_0 = A_0 + iB_0 .$$

The weights on the second hub have the center of mass

$$c_1 r = A_1 + iB_1 \qquad \text{etc.},$$

and on the last hub,

$$c_n r^n = A_n + iB_n$$

wherein the positive real ξ axis is chosen downward, the positive imaginary axis to the right, etc. The center of mass of the whole system naturally assumes the lowest position. The system rolls along on the knife edges until that is the case, i.e., until the angle of the positive ξ axis is $+\Phi$ with respect to the vertical. By means of the worm gear, the individual hubs are then rotated through the angles φ, 2φ, $\cdots$, $n\varphi$ with respect to the first. Then the Φ of the series can be read off for the same r and various values of φ. The best method of operation is to seek values of φ by a continuous variation, for which the positive ξ axis forms the angles $\Phi = 0$, $+\pi/4$, $+\pi/2$, $\cdots$ with the vertical, successively. The corresponding values of φ are then read off, and the corresponding value of ξ is plotted in the r, φ plane at this point. If this is carried out for several values of r, the points of the lines $\Phi = c$ are obtained on concentric circles. Connection of these points gives an approximate drawing of the lines $\Phi = c$, and the approximate values of the roots can then be found.

For more precise evaluation of such a root, the equation can be developed by Horner's method about the approximation value. Then, by mechanical means, the lines $\Phi = c$ can be determined on concentric circles about this point, or the value can be improved by Newton's method (cf. 11).

8. Runge[7] has given *a graphical method for the approximate construction of the complex roots of a rational integral function with complex or real roots,*

$$g(z) = c_n z^n + c_{n-1} z^{n-1} + \cdots + c_1 z + c_0 . \tag{11}$$

He generalized the method of Lill (7.6; 9.5). Just as in this latter method, the coefficients

$$c_n , \quad c_{n-1} , \quad c_{n-2} \cdots c_2 , \quad c_1 , \quad c_0 \tag{11a}$$

are added vectorially in the complex plane, and the framework representing the function is obtained in the polygon $OC_nC_{n-1} \cdots C_0$.

To construct the value of the function for $z = z_0$, we draw the vector $C_nD_{n-1} = c_nz_0$, obtained by multiplication of the two vectors c_n and z_0 . Therefore

$$C_{n-1}D_{n-1} = d_{n-1} = c_nz_0 + c_{n-1} .$$

If we carry out the construction in the simplest way by subtraction of the phase angles and by calculation with a slide rule of the adjacent sides of a triangle $C_{n-1}D_{n-1}D_{n-2}$ similar to C_nOD_{n-1}, then

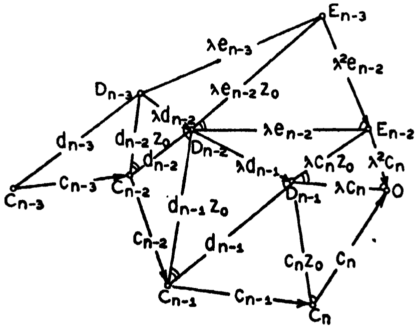


FIG. 86

$$C_{n-1}D_{n-2} = d_{n-1}z_0 = c_nz_0^2 + c_{n-1}z_0$$

$$C_{n-2}D_{n-2} = c_nz_0^2 + c_{n-1}z_0 + c_{n-2} = d_{n-2} \text{ etc.}$$

In this step we describe circles with the same radii about all points C, so that we have only to draw in equal chords with a compass. The last length is

$$C_0D_0 = d_0 = d_1z_0 + c_0 = c_nz_0^n + c_{n-1}z_0^{n-1} + \cdots + c_1z + c_0 = g(z_0).$$

As we have seen in the consideration of the real case in 7.1:

$$g(z) = g(z_0) + (z - z_0)(d_1 + d_2z + d_3z^2 + \cdots + d_{n-1}z^{n-2} + c_nz^{n-1})$$

$$= g(z_0) + [zz_0](z - z_0).$$

This can also be verified by multiplication, keeping in mind the value of d. The framework of the divided difference of the first order, or of the first different quotient (8.1)

$$(12) \quad [zz_0] = g_1(z) = \frac{g(z) - g(z_0)}{z - z_0} = d_1 + d_2z + \cdots + d_{n-1}z^{n-2} + c_nz^{n-1}$$

is now represented, with another scale of measurement, by the line sequence $D_0D_1 \cdots D_{n-1}O$. If the ratio OD_{n-1}/OC_n is denoted by λ, then, by similar triangles,

$$(12a) \qquad D_{n-1}O = \lambda c_n \, ; \, D_{n-2}D_{n-1} = \lambda d_{n-1} \cdots D_1D_0 = \lambda d_0 \, .$$

We can again expand a line sequence $OE_{n-2}E_{n-3} \cdots E_0$ over this framework with the same value z_0, while we have

$$\Delta OC_nD_{n-1} \sim \Delta OD_{n-1}E_{n-2} \sim \Delta E_{n-2}D_{n-2}E_{n-3} \cdots, \text{ etc.} \tag{12b}$$

In this case,

$$\begin{aligned} D_{n-1}E_{n-2} &= \lambda c_n z_0 , \\ D_{n-2}E_{n-2} &= \lambda(c_n z_0 + d_{n-1}) = \lambda e_{n-2} \\ D_{n-2}E_{n-3} &= \lambda(c_n z_0^2 + d_{n-1}z_0), \\ D_{n-3}E_{n-3} &= \lambda(c_n z_0^2 + d_{n-1}z_0 + d_{n-2}) = \lambda e_{n-2} \end{aligned} \tag{12c}$$

etc., i.e., since

$$\begin{aligned} g(z) &= g(z_0) + g_1(z_0)(z - z_0) + (z - z_0)^2(e_1 + e_2 z + \cdots \\ &\qquad + e_{n-2}z^{n-3} + c_n z^{n-2}) \\ &= g(z_0) + [z_0z_0](z - z_0) + [zz_0z_0](z - z_0)^2. \end{aligned} \tag{13}$$

The line sequence $E_0E_1E_2 \cdots E_{n-2}0$ is then the framework of the function appearing in the final parenthesis, and we have

$$E_{n-2}O = \lambda^2 c_n , \qquad E_{n-3}E_{n-2} = \lambda^2 e_{n-2} \cdots . \tag{13a}$$

If we proceed in this fashion, the coefficients of the development of $g(z)$ by powers of $z - z_0$ can be found graphically. These are

$$g(z_0) = d_0 = C_0D_0 , \ g_1(z_0) = e_0 = \frac{D_0E_0}{\lambda}, \ g_2(z_0) = f_0 = \frac{E_0F_0}{\lambda^2}, \tag{13b}$$

etc. Naturally the coefficients of a product development can be found in an entirely similar manner.

9. If a root of such an equation is to be determined, then we first seek a value z_1 by trial, for which the point representing the function value $g(z_1)$ lies in the neighborhood of C_0. Then we seek a second value $\bar{z}_1$ which gives a point $\overline{D}_0$ for the function value $g(\bar{z}_1)$, which also lies in the vicinity of C_0. Now by the method of false position for complex values (18.3), we get a better value z_2 if we make $\Delta D_0\overline{D}_0C_0 \sim \Delta z_1\bar{z}_1z_2$, since the neighborhood of a simple root is mapped conformally to the neighborhood of C_0.

But we can also get a better approximation by a transference of New-

ton's approximation method for complex roots into graphical form. If $\bar{z}_1$ lies in the neighborhood of the value of the root z_0, then we can terminate the above development with the linear terms. Then

$$g(\bar{z}_1) - g(z_0) = g_1(z_1)(\bar{z}_1 - z_0)$$

or in vector notation, since $\overline{D}_0$ coincides with C_0,

$$\frac{C_0D_0}{D_0E_0} = \frac{\bar{z}_0 - z_0}{\lambda} = \frac{c_n\bar{z}_0 - c_nz_0}{\lambda c_n} = \frac{\overline{D}_{n-1}D_{n-1}}{D_{n-1}O},$$

i.e.,

$$\Delta D_0C_0E_0 \sim \Delta D_{n-1}\overline{D}_{n-1}O.$$

Therefore, if we have found a polygon $O, D_{n-1}, \cdots, D_0$, whose endpoints lie in the neighborhood of C_0, then the polygon $O, E_{n-2}, E_{n-3}, \cdots, E_0$, is constructed to the same value z_0. Therefore a point $\overline{D}_{n-1}$ is obtained for which the corresponding $z = C_n\overline{D}_{n-1} : OC_n$ gives a better approximation to the value of the root, if the two triangles have the relation $C_0D_0E_0 \sim \overline{D}_{n-1}D_{n-1}O$, as is shown in Fig. 87.

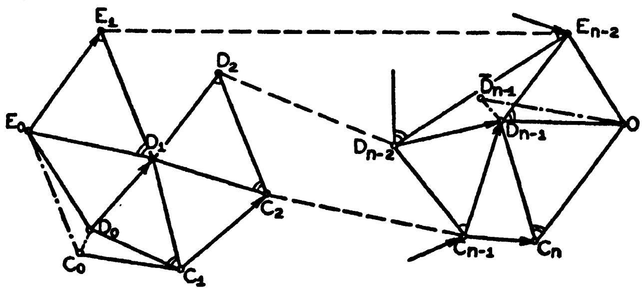


FIG. 87

If a value of the root z_0 is found, then all other roots are roots of $g_1(z)$. The other roots will therefore be determined on the polygon $C_0, D_1, \cdots, \cdots, D_{n-1}O$ as a framework. If the second root has also been found, then the other roots are also roots of $g_2(z)$, and we seek to construct the next root on the polygon $C_0, E_1, E_2, \cdots, E_{n-2}O$, etc. We must therefore keep in mind that the points with the index zero all lie near C_0. But, since the inaccuracy of each root affects the construction of the next, it is advisable to improve the approximation values thus found. This should be done graphically on the framework of the equation originally given, or numerically by iteration or Newton's method. The details are illustrated in the following example.

10. *Example*: The roots of the equation

$$z^3 + (1 - i)z^2 - (1 + 3i)z + 3 + 2i = 0$$

are to be found. The coefficients give the polygon $C_0C_1C_2C_3O$. To be able to draw the similar triangles conveniently, circles are described about the points C_1, C_2, C_3, O with radii of 2-3 cm., so that the chords can be drawn with a compass. The ratio $|D_2O| : |C_3O|$ is calculated with a slide rule. We then measure the side C_2D_2, and read off the corresponding length C_2D_1 on the slide rule. We measure off this length on the free side of the angle OC_3D_2 the angle having been measured by the compass C_2D_2. Hence the leg does not need to be drawn. The mark of the point D is sufficient. A first try gives (Fig. 88) points $D_2'D_1'D_0'$. In this simple case, we can, with some considera-

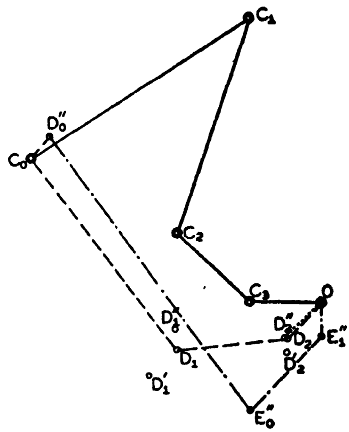


FIG. 88

tion, find a better point D_2'', which gives the points $D_1''D_0''$. The latter point lies sufficiently close to C_0 for us to construct the line sequence $E_1''E_0''$ also. The new approximation value then lies in the same relation to OD_2'' as C_0 to $E_0''D_0''$. The new vector polygon $OD_2D_1C_0$ actually leads to the point C_0, within the accuracy of the drawing. From this we get as the approximate root

$$z_1 = \frac{C_3D_2}{C_3O} = 0.56 - 0.52i.$$

The other two roots are also the roots of the equation whose coefficients are proportional to the vectors of the polygon $C_0D_1D_2O$. This line sequence is drawn in Fig. 89. The roots of the equation

$$g_1(z) = c_3z^2 + d_1z + d_0 = 0$$

can be constructed with ruler and compass. If we divide D_2D_1 at R, then $D_2R : D_2O = -d_1 : 2c_3 = p$, and if the new unknown $u = z - p$ is introduced, the above equation becomes

$$c_3u^2 + g_1(p) = 0.$$

The value $g_1(p)$ can be constructed from the polygon ORR_1; it is equal to C_0R_1. The framework of the equation for u can be constructed from this; it is the polygon $R_2'RO'$ where $R_2'R = C_0R_1$ and $RO' = D_2O$

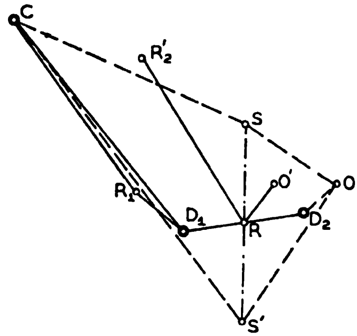


FIG. 89

The endpoint S of the solution polygon must be so placed that $\Delta O'RS \sim \Delta SRR_2'$, i.e., S must lie on the bisector of the angle $O'RR_2'$. In addition, $|RS|$ must be the mean proportional between the lengths $|O'R|$ and $|RR_2'|$. We then get the two values of u:

$$u_1 = RS/RO', \qquad u_2 = RS'/RO',$$

and the value of z from the dotted solution polygon:

$$z_2 = D_2S/D_2O = \frac{-0.71 + 1.22i}{0.44 + 0.52i} = 0.69 + 1.95i$$

$$z_3 = D_2S'/D_2O = \frac{-0.82 - 1.33i}{0.44 + 0.52i} = -2.27 - 0.343i.$$

In Fig. 89, the vector polygon is intentionally drawn in the original position. It is more practical to take D_2O as the real axis and OD_2 as the scale modulus.

Lill's method, mentioned in 7.6 and 19.5, is contained as a special case of the above. If a new unknown $z = iv$ is introduced in the equation

$$g(z) = a_nz^n + a_{n-1}z^{n-1} + \cdots + a_1z + a_0,$$

then, under the assumption that all coefficients a are real, the vectors of the equation framework become

$$a_n, \qquad -ia_{n-1}, \qquad -a_{n-2}, \qquad +ia_{n-3}, \qquad a_{n-4}, \qquad \cdots,$$

i.e., each vector is rotated 90° with respect to the preceding one. This rotation is counterclockwise, if we begin with a_n, as in 7.6. For the real

roots, $v = ix$ becomes purely imaginary so that the points $D_{n-1} \cdots D_0$ lie on the polygon itself, and also form a rectangular polygon.

In equations with real coefficients, the roots either are real or are complex conjugates. In this case, the construction is simplified, as is shown in the original work.

11. The approximation values thus found can now be improved by Newton's method for one variable, because the proof of the convergence in 18.11 holds without any assumption on the real nature of the functions or variables. The process converges by 18(27) if

$$\left| \frac{g(z) \cdot g''(z)}{[g'(z)]^2} \right| < 1.$$

Horner's scheme can be used for the calculation of $g(z)$ and $g'(z)$, e.g., for the first root of the example in Sec. 10,

1	$1 \quad i$	$-f - 3i$	$3 + 2i$
	$0.56 - 0.52i$	$0.874 - 0.811i$	$-0.513 + 0.476i$
		$-0.790 - 0.851i$	$-2.424 - 2.613i$
1	$1.56 - 1.52i$	$-0.916 - 4.662i$	$+0.063 - 0.135i = g(z)$
	$0.56 - 0.52i$	$+1.187 - 1.102i$	
		$-1.061 - 1.142i$	
1	$2.32 - 2.04i$	$-0.790 - 6.906i = g'(z)$	
	$0.56 - 0.52i$		
1	$2.68 - 2.56i = g''(z)/2.$		

Here

$$m = \left| \frac{(0.063 - 0.135i)(2.68 - 2.56i) \cdot 2}{(-0.790 - 6.906i)^2} \right|$$

$$= \left| \frac{-0.354 - 1.046i}{47.069 + 10.911i} \right| = 0.023.$$

The process is then certainly convergent and

$$\Delta z = + \frac{(0.063 - 0.135i)(0.790 - 6.906i)}{48.317} = -0.0183 - 0.0112i.$$

Therefore $\bar{z} = 0.542 - 0.531i$. The value of the function belonging to this argument is

$$g(\bar{z}) = +0.0020 - 0.0006i.$$

If this is not sufficient, the same calculation can be performed a second time. A second correction is then obtained

$$\Delta\bar{z} = -0.00005 - 0.00030i.$$

NOTES

1. Deimler, *Z. f. Flugtechnik und Motorluftschiffahrt* I (1910).

2. Hurwitz, *Math. Annalen*, **46** (1895), pp. 273-284. Schur, *Z. f. angew. Math. u. Mech.* I (1921), pp. 307-311.

3. Thommeck, *Über einen mechanischen Apparat zur Bestimmung der Nullstellen ganzer rationaler Funktionen mit komplexen Koeffizienten*. Dissertation (Bonn, 1917).

4. Runge, *Göttinger Nachrichten Math. phys. Klasse* (1917).

22. Graeffe's Method.

1. While the methods mentioned so far permit us to calculate only one of the roots of an equation, a method has been given independently by Dandelin,[1] Lobachevsky,[2] and Graeffe[3] for algebraic equations with real coefficients, which yields *approximate values for all real and complex roots simultaneously*.

The method can be analyzed most simply for the case that *all the roots* of the equation

$$g(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0 = 0 \tag{1}$$

are real, and are different from one another in absolute value. Then

$$|x_1| > |x_2| > |x_3| \cdots > |x_n|. \tag{2}$$

The problem is to produce a new equation from the given equations, whose roots are the squares of the roots of the given equation. From this we form a new equation, whose roots are the squares of the roots of the equation just obtained and hence the fourth power of the roots of the given equation, etc. The difference between the individual powers of the roots of the original equation naturally becomes greater with the increase of the exponent. Finally, if we proceed with the method given above, we get an equation

$$r_n u^n + r_{n-1} u^{n-1} + \cdots + r_2 u^2 + r_1 u + r_0 = 0 \tag{3}$$

in which the difference between the roots is so great that $-r_{n-1}/r_n$ should equal the sum of the roots, by Vieta's root theorem, and differs from the largest root only by a very small fraction of its value. In the same way, r_{n-2}/r_n, which is equal to the sum of all possible products of pairs of roots, differs from the product of the two largest roots by a relatively small value, etc. With further approximation, we can set

$$u_1 = -\frac{r_{n-1}}{r_n}; \qquad u_1 \cdot u_2 = \frac{r_{n-2}}{r_n}; \qquad u_1 \cdot u_2 \cdot u_3 = -\frac{r_{n-3}}{r_n}; \qquad \cdots \tag{4}$$

or

$$u_1 = -\frac{r_{n-1}}{r_n}; \qquad u_2 = -\frac{r_{n-2}}{r_{n-1}}; \qquad u_3 = -\frac{r_{n-3}}{r_{n-2}}; \qquad \cdots . \tag{5}$$

The given equation is then broken up into n equations

$$(6)\qquad r_n u + r_{n-1} = 0; \quad r_{n-1}u + r_{n-2} = 0; \quad r_{n-2}u + r_{n-3} = 0; \quad \cdots$$

whose roots are known powers of the roots of the given equation, with good approximation. The desired roots can then be obtained from these powers, except for the sign.

2. To form the *equation whose roots are the squares of the roots of the given equation,* we must observe that if we replace x by $-x$ in the given equation, the roots of this new equation are equal in absolute value to the roots of the given equation. The signs of the roots, however, are opposite. In the resultant equation, the coefficients of the odd powers all have opposite signs to those in the given equation. We have therefore the two equations

$$(7)\qquad \begin{aligned} a_n x^n + a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \cdots + a_2x^2 + a_1x + a_0 &= 0, \\ a_n x^n - a_{n-1}x^{n-1} + a_{n-2}x^{n-2} - \cdots \pm a_2x^2 \mp a_1x \pm a_0 &= 0, \end{aligned}$$

or, if we express the two equations in products,

$$(8)\qquad \begin{aligned} a_n(x - x_1)(x - x_2), \cdots, (x - x_n) &= 0, \\ a_n(x + x_1)(x + x_2), \cdots, (x + x_n) &= 0. \end{aligned}$$

If these two equations are multiplied together, we obtain an equation of nth degree in $z(z = x^2)$:

$$(9)\qquad a_n^2(z - x_1^2)(z - x_2^2)(z - x_3^2) \cdots (z - x_n^2) = 0.$$

The roots of this equation are equal to the squares of the roots of the given equation. The multiplication is best carried out in the form of a table which is written out here for the case of the equation of fourth degree:

		x^4	x^3	x^2	x	
1.	I Equation	a_4	a_3	a_2	a_1	a_0
	II Equation	a_4	$-a_3$	a_2	$-a_1$	a_0
		a_4^2	$-a_3^2$	a_2^2	$-a_1^2$	$+a_0^2$
			$+2a_2a_4$	$-2a_1a_3$	$+2a_0a_2$	
				$+2a_0a_4$		
2.	I Equation	b_4	$-b_3$	b_2	$-b_1$	b_0
	II Equation	b_4	$+b_3$	b_2	$+b_1$	b_0
		b_4^2	$-b_3^2$	$\cdots$		

The b's are the sums of the vertical columns; b_4 is the coefficient of $z^4 = x^8$, $-b_3$ of $z^3 = x^6$, etc. In equations of higher degree, the number of terms in a column becomes correspondingly greater. The b's are the coefficients of the equation whose roots are $x_1^2, x_2^2, \cdots, x_n^2$. Since the roots of this equation are even powers of the corresponding roots of the original equation, they are certainly positive, so that the 2nd, 4th, $\cdots$ coefficients of the equation must be negative. We proceed with this equation as with the first, etc. The details are best illustrated by the example below.

3. *Example*: We seek the roots of the equation

$$g(x) = x^3 + 5x^2 - 3x - 1 = 0.$$

The following table, which is prepared with the use of four place tables of squares, is self-explanatory. Instead of writing four place numbers, the so-called normal values are written (Schneider, 1825); i.e., the first digit is followed by a decimal and the exponent of the power of ten with which this normal value is to be multiplied is written as a superscript.

1.	I.	1	$+5$	-3	-1
	II.	1	-5	-3	$+1$
		1	-2.5^1	$+0.9^1$	-1
			-0.6^1	$+1^1$	
2.	I.	1	-3.1^1	$+1.9^1$	-1
	II.	1	$+3.1^1$	$+1.9^1$	$+1$
		1	-9.61^2	$+3.61^2$	-1
			$+0.38^2$	-0.62^2	
4.	I.	1	-9.23^2	$+2.99^2$	-1
	II.	1	$+9.23^2$	$+2.99^2$	$+1$
		1	-8.519^5	$+8.940^4$	-1
			$+0.006^5$	-0.185^4	
8.	I.	1	-8.513^5	$+8.755^4$	-1
	II.	1	$+8.513^5$	$+8.755^4$	$+1$
		1	-7.247^{11}	$+7.663^9$	-1
			$\cdots$	$+0.002^9$	
		1	-7.247^{11}	$+7.663^9$	-1 .

The doubled products always become smaller and can be neglected in the calculation of the equation which has as roots the sixteenth powers of the roots of the given equation. The criterion of this is

that within the digits written out, the summands, except those which contain the largest root, are negligible. This is easily seen from the following consideration.

If, for example, the equation is of the 8th power,

$$d_3z^3 - d_2z^2 + d_1z - d_0 = 0, \tag{10}$$

the approximate values for the roots, calculated by Graeffe's method, are

$$x_1 = \left(+\frac{d_2}{d_3}\right)^{1/8}; \qquad x_2 = \left(+\frac{d_1}{d_2}\right)^{1/8}; \qquad x_3 = \left(+\frac{d_0}{d_1}\right)^{1/8}. \tag{11}$$

If we multiply this equation by

$$d_3z^3 + d_2z^2 + d_1z + d_0 = 0 \tag{12}$$

to get the equation for the sixteenth power, then we get

$$d_2^2u^3 - d_2^2u^2 + d_1^2u - d_0^2 = 0 \tag{13}$$

since the doubled products no longer affect the digits shown. From this equation we obtain the approximation values for the roots of the given equation:

$$x_1 = \left(\frac{d_2^2}{d_3^2}\right)^{1/16}; \qquad x_2 = \left(\frac{d_1^2}{d_2^2}\right)^{1/16}; \qquad x_3 = \left(\frac{d_0^2}{d_1^2}\right)^{1/16}. \tag{14}$$

These are the same values which are obtained from the equation for the 8th power. Therefore, if the doubled products no longer affect the written digits of the squares in the first series of the columns, there is no point in carrying the calculations any further.

The same scheme of calculation is used if we want to carry through the calculation with a slide rule or calculating machine. When logarithms are used, we employ a similar arrangement, in which the logarithms are written in place of the numbers, and the addition or subtraction of two logarithms is carried out with Gaussian logarithms.

4. The *calculation of the roots* in the above example with four place logarithms is most easily done by the following scheme:

	Δ	$\Delta:16$	
$\log 1 = 0.0000$			
	$0.1156 - 10$	$0.3822 - 1$	$\lvert x_3 \rvert = 0.2411$
$\log 7.663^9 = 9.8844$			
	$0.0243 - 2$	$0.8765 - 1$	$\lvert x_2 \rvert = 0.7525$
$\log 7.247^{11} = 11.8601$			
	11.860	0.7413	$\lvert x_1 \rvert = 5.511.$
$\log 1 = 0.0000$			

A separate calculation is required *to determine the signs.* It is simplest to calculate the value of the function corresponding to the equation for several values of x which lie between the roots. We find

$$g(-\infty) = -\infty, \qquad g(-1) = +6, \qquad g\left(-\frac{1}{2}\right) = +\frac{13}{8},$$

$$g(0) = -1, \qquad g(+1) = +2, \qquad g(+\infty) = +\infty.$$

Therefore one root lies between $-\infty$ and -1, a second between -0.5 and 0, and the third between 0 and $+1$. This is only possible if

$$x_1 = -5.511, \qquad x_2 = +0.7525, \qquad x_3 = -0.2411.$$

For proof of the accuracy obtained, we can calculate the corrections to be added to the roots, by means of Horner's scheme (cf. 19.1). We find

$$\Delta x_1 = +0.0004; \qquad \Delta x_2 = +0.000013; \qquad \Delta x_3 = -0.000011.$$

The corrections therefore affect only the fourth decimal place.

5. The course of calculation is somewhat different if the equation has *groups of roots of nearly the same absolute value* or pairs of complex conjugate roots whose absolute values are of course equal.[4] A complete separation into linear equations is not possible by the operation described above. In the second case, for example, the complex roots are determined from quadratic equations, while linear equations are again obtained for single roots.

6. The general proof will not be given here. Instead, we will work through one *example.* We take the equation

$$x^4 - 4x^3 + 5x^2 + 5x - 10 = 0$$

and use the calculation scheme given in Sec. 2:

1.	I.	1	-4	$+5$	$+5$	-10
	II.	1	$+4$	$+5$	-5	-10
		1	-1.6^{1}	$+2.5^{1}$	-2.5^{1}	$+1^{2}$
			$+1^{1}$	$+4^{1}$	-10^{1}	
				-2^{1}		
2.	I.	1	-0.6^{1}	$+4.5^{1}$	-1.25^{2}	$+1^{2}$
	II.	1	$+0.6^{1}$	$+4.5^{1}$	$+1.25^{2}$	$+1^{2}$
		1	-3.6^{1}	$+2.025^{3}$	-1.563^{4}	1^{4}
			$+9^{1}$	-1.5^{3}	$+0.9^{4}$	
				$+0.2^{3}$		

4.	I.	1	$+5.4^{1}$	$+7.25^{2}$	-6.63^{3}	$+1^{4}$
	II.	1	-5.4^{1}	$+7.25^{2}$	$+6.63^{3}$	$+1^{4}$
		1	-2.916^{3}	$+5.256^{5}$	-4.396^{7}	$+1^{8}$
			$+1.45^{3}$	$+7.160^{5}$	$+1.45^{7}$	
				$+0.2^{5}$		
8.	I.	1	-1.466^{3}	$+1.262^{6}$	-2.946^{7}	$+1^{8}$
	II.	1	$+1.466^{3}$	$+1.262^{6}$	$+2.946^{7}$	$+1^{8}$
		1	-2.149^{6}	$+1.593^{12}$	-8.679^{14}	1^{16}
			$+2.524^{6}$	-0.086^{12}	$+2.524^{14}$	
				...		
16.	I.	1	3.75^{5}	$+1.507^{12}$	-6.155^{14}	1^{16}
	II.	1	-3.75^{5}	$+1.507^{12}$	$+6.155^{14}$	1^{16}
		1	-1.406^{11}	$+2.271^{24}$	-3.788^{29}	1^{32}
			$+30.14^{11}$	...	$+0.301^{29}$	
				...		
32.	I.	1	2.873^{12}	$+2.271^{24}$	$+3.487^{29}$	1^{32}
	II.	1	-2.873^{12}	$+2.271^{24}$	-3.487^{29}	1^{32}
		1	-8.254^{24}	$+5.158^{48}$	-1.216^{59}	1^{64}
			$+4.542^{24}$	...	$+0.005^{59}$	
				...		
64.		1	-3.712^{24}	$+5.158^{48}$	-1.211^{59}	1^{64}.

In the equation whose roots are the 32nd powers of the roots of the given equation, the influence of the products on the last three terms is already very small. Hence in a further development of the scheme they vanish completely. In the second term, on the other hand, no decrease is observable in the influence of the products. Now the equation has two complex roots, as can be determined by the rules given in earlier articles. Therefore a further development of the scheme is unnecessary. We now calculate the absolute value of the real roots by use of four place logarithms, according to the scheme given in Sec. 4, and also the value of the constant term of the quadratic equation, which is equal to the product of the two complex conjugates.

	Δ	$\Delta:64$	
$\log 1^{64} = 64.0000$			
	4.9168	0.0768	$\lvert x_1 \rvert = 1.194$
$\log 1.211^{59} = 59.0832$			
	10.3708	0.1620	$\lvert x_2 \rvert = 1.452$
$\log 5.158^{48} = 48.7124$			
	48.7124	0.7611	$r^2 = 5.769.$
$\log 1 = 0.0000$			

To determine the signs of the two real roots, we calculate the value of the function, whose roots we seek, for a value of x which lies between their absolute values, for example, $x = 1.3$:

1	−4	+5	+5	−10
	1.3	−3.51	+1.94	+ 9.02
	−2.7	+1.49	+6.94	− 0.98.

Since the function is positive for $x = -\infty$ and $x = +\infty$, but negative for $x = 0$, the smaller root must be negative, the larger positive, i.e., $x_1 = -1.194$; $x_2 = +1.452$.

Now the factor of x^3 in the equation is equal to the negative sum of the roots. Therefore, if the two complex roots are $u \pm iv$, then

$$x_1 + x_2 + 2u = 4 = 0.258 + 2u, \qquad u = 1.871.$$

We calculate v from

$$v = (r^2 - u^2)^{1/2} = (5.769 - 3.501)^{1/2} = 1.506.$$

The two complex roots are then

$$x_{3,4} = 1.871 \pm 1.506i.$$

If these values are corrected by Newton's approximation method (19.1 and 21.11), we get as the new approximation values

$$x_1 = -1.19402, \qquad x_2 = 1.45179, \qquad x_{3,4} = 1.87113 \pm 1.50589i.$$

If $v = 0$, then $u = r$, so that there are two equal roots (or two with opposite signs).

A thorough treatment of the Graeffe method is given by Encke.[5]

Polya[6] has shown that this can also be used for the approximate calculation of the roots of any power series. We shall not discuss this here. Likewise we shall only mention other methods for the approximate calculation of the roots of equations, such as the method of Bernoulli[7] for the calculation of the roots with the largest and smallest absolute values, that of Jacobi[8] in the case that several roots are present with the same maximum absolute value, that of Lagrange[9] which gives the approximation value in the form of a continued fraction, also the method of Laguerre[10] which is to be developed with invariant series methods, and finally the formula of Whittaker[11] which gives a series development for the smallest root, whose individual terms are quotients of determinants. These determinants are formed by the coefficients of the equations.

NOTES

1. Dandelin, *Mém. de. l'Acad. Royale de Bruxelles*, **3** (1826), p. 48.
2. Lobachevsky, *Algebra* (Kazan, 1834), Art. 257.
3. Graeffe, *Auflösung der höheren numerischen Gleichungen* (Zürich, 1837).
4. For example, Runge, *Praxis der Gleichungen*, 1st ed. (Leipzig, 1900), p. 163 ff.
5. Encke, *Berliner Astronomisches Jahrbuch* (1841). *Ges. Abh. I* (Berlin, 1888), pp. 125-187.
6. Polya, *Z. f. Math. u. Phys.* **63** (1914-15), p. 275.
7. Bernoulli, *Commentarii Acad. Petropol.* **3** (1732); Euler, *Introductio in Analysin Inf. I* (Lausanne, 1748), Ch. XVII.
8. Jacobi, *Journal für Math.* **13** (1835), p. 340.
9. Larange, *De la résolution des équations numériques de tous les dégrés* (Paris, 1798), Ch. III.
10. Laguerre, *Annales de math.*, series 2, **19** (1880). *Oeuvres I* (Paris, 1898), p. 87.
11. Whittaker, *Proc. Edin. Math. Soc.* **36** (1918), p. 103.

23. Linear Equations with Several Unknowns.

1. There are two important methods for the numerical solution of n *linear equations with n unknowns*; first, the *subtraction method* in the form which Gauss gave for the solution of the normal equations in the calculation of errors by the method of least squares, and second, the *substitution method*, in which the unknown from one equation is expressed as a function of the others and substitution is then made in the other equations. One root after another is then eliminated, i.e., n equations with n unknowns are converted to $n - 1$ equations with $n - 1$ unknowns, these in turn become $n - 2$ equations with $n - 2$ unknowns, etc. Since this method is the same for all cases, we will only develop the case of three equations with three unknowns. We write the equations in the form

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + a_1 &= 0, \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + a_2 &= 0, \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + a_3 &= 0. \end{aligned} \tag{1}$$

If the first equation is multiplied with a_{21}/a_{11} (this can be done with one setting of a slide rule, or a single division with the calculating machine, setting up the quotient in the set up mechanism and multiplying by the separate coefficients of the first equation), and the result subtracted from the second equation, and if the first equation is also multiplied by a_{31}/a_{11} and the result subtracted from the third equation, then the terms in x drop out, and there remains

$$\left(a_{22} - \frac{a_{21}}{a_{11}} a_{12}\right)x_2 + \left(a_{23} - \frac{a_{21}}{a_{11}} a_{13}\right)x_3 + \left(a_2 - \frac{a_{21}}{a_{11}} a_1\right) = 0,$$

(2)

$$\left(a_{32} - \frac{a_{31}}{a_{11}} a_{12}\right)x_2 + \left(a_{33} - \frac{a_{31}}{a_{11}} a_{13}\right)x_3 + \left(a_3 - \frac{a_{31}}{a_{11}} a_1\right) = 0.$$

If we could cancel subscripts, the two summands in each bracket would have the same indices. For abbreviation then, we write the equations (2) in the form

(3)

$$a'_{22}x_2 + a'_{23}x_3 + a'_2 = 0,$$

$$a'_{32}x_2 + a'_{33}x_3 + a'_3 = 0.$$

If the first equation is now multiplied by a'_{32}/a'_{22} and the result subtracted from the first equation, then

(4)
$$\left(a'_{33} - \frac{a'_{32}}{a'_{22}} a'_{23}\right)x_3 + \left(a'_3 - \frac{a'_{32}}{a'_{22}} a'_2\right) = 0,$$

where, after the second elimination, we place two primes on the coefficients:

(5)
$$a''_{33}x_3 + a''_3 = 0.$$

The value x_3 is calculated from this equation. The other two unknowns can be calculated by substitution. For this purpose, an equation is chosen from each group, e.g.,

(6)

$$a''_{33}x_3 + a''_3 = 0,$$

$$a'_{23}x_3 + a'_{22}x_2 + a'_2 = 0,$$

$$a_{13}x_3 + a_{12}x_2 + a_{11}x_1 + a_1 = 0.$$

This is a *reduced equation system*, from which, starting with the first equation, the individual unknowns can be calculated, one after the other. We also substitute in all the equations of the group so as to have a check on the accuracy.

Possible errors of calculation can also be avoided if the second and third terms of the given equation are exchanged in the calculation of x_2, and the first and third in the calculation of x_1. The same elimination process is then carried out as above.

A scheme for the above operations would appear as follows, where unnecessary writing of the variables is avoided:

x_1	x_2	x_3		s	
a_{11}	a_{12}	a_{13}	a_1	s_1	a_1
a_{21}	a_{22}	a_{23}	a_2	s_2	$a_{13}x_3$
	$\frac{a_{21}}{a_{11}}\,a_{12}$	$\frac{a_{21}}{a_{11}}\,a_{13}$	$\frac{a_{21}}{a_{11}}\,a_1$	$\frac{a_{21}}{a_{11}}\,s_1$	$a_{12}x_2$
					a_1'
a_{31}	a_{32}	a_{33}	a_3	s_3	
	$\frac{a_{31}}{a_{11}}\,a_{12}$	$\frac{a_{31}}{a_{11}}\,a_{13}$	$\frac{a_{31}}{a_{11}}\,a_1$	$\frac{a_{31}}{a_{11}}\,s_1$	$x_1 = -\frac{a_1'}{a_{11}}$
	a_{22}'	a_{23}'	a_2'	s_2'	a_2'
					$a_{23}'x_3$
	a_{32}'	a_{33}'	a_3'	s_3'	
					a_2''
		$\frac{a_{32}'}{a_{22}'}\,a_{23}'$	$\frac{a_{32}'}{a_{22}'}\,a_2'$	$\frac{a_{32}'}{a_{22}'}\,s_2'$	$x_2 = -\frac{a_2''}{a_{22}'}$
		a_{33}''	a_3''	s_2''	$x_3 = -\frac{a_3''}{a_{33}''}$

The s column serves as a control. The value s_n should be the sum of the coefficients of the nth equation. It is easily verified that if the same operations are undertaken with these sums as with the corresponding equations, the sum of the coefficients of the above equations is always obtained, e.g.,

$$
\begin{aligned}
s_2' = s_2 - \frac{a_{21}}{a_{11}}\,s_1 &= \left(a_{21} - \frac{a_{21}}{a_{11}}\,a_{11}\right) + \left(a_{22} - \frac{a_{21}}{a_{11}}\,a_{12}\right) \\
&+ \left(a_{23} - \frac{a_{21}}{a_{11}}\,a_{13}\right) + \left(a_2 - \frac{a_{21}}{a_{11}}\,a_1\right) = 0 + a_{22}' + a_{23}' + a_2' \,.
\end{aligned} \tag{7}
$$

If the calculation operations for the sums are carried out in the above table, then a rather certain check is had on the calculations. The scheme can be shortened somewhat if we omit the quantities a_{33}', a_3', etc., and form $a_{33} - (a_{31}/a_{11})\,a_{13} - (a_{32}'/a_{22}')\,a_{23}'$ directly. Still, the certainty of the calculation suffers if such is done.[1]

2. *Example*: A tractive force of one ton pulls in the direction ST on a triple jack, whose point lies 6 m. above the surface, and the base

points of which have the positions shown in Fig. 90. How great are the tensions in the different members?

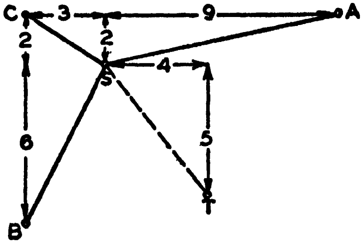


FIG. 90

The rod lengths are $AS = 11, BS = 9, CS = 7$, and the cable length ST is $(77)^{1/2}$. If the tensions in the beams are denoted by ζ_1, ζ_2 and ζ_3, and the equilibrium conditions for the components in the three directions of the coordinate axes are written down, we obtain the three equations

$$x \text{ Direction:}\quad \frac{9}{11}\zeta_1 - \frac{3}{9}\zeta_2 - \frac{3}{7}\zeta_3 + \frac{4}{(77)^{1/2}}$$

$$= \quad 0.818\zeta_1 - 0.333\zeta_2 - 0.429\zeta_3 + 0.456 = 0,$$

$$n \text{ Direction:}\quad -\frac{2}{11}\zeta_1 + \frac{6}{9}\zeta_2 - \frac{2}{7}\zeta_3 + \frac{5}{(77)^{1/2}}$$

$$= -0.182\zeta_1 + 0.667\zeta_2 - 0.286\zeta_3 + 0.570 = 0.$$

$$y \text{ Direction:}\quad +\frac{6}{11}\zeta_1 + \frac{6}{9}\zeta_2 + \frac{6}{7}\zeta_3 + \frac{6}{(77)^{1/2}}$$

$$= +0.545\zeta_1 + 0.667\zeta_2 + 0.857\zeta_3 + 0.684 = 0.$$

In this case, a tension in the beams is taken as positive. If we multiply by 1000 to save writing, the scheme becomes

	ζ_1	ζ_2	ζ_3		s	
I	818	−333	−429	456	512	456
II	−182	667	−286	570	769	−156.5
Ia		+ 74	+ 96	−101	−113	+298.4
III	+545	667	857	684	2753	597.9
Ib		−222	−286	+304	341	$\zeta_1 = -0.731$ (Compression)
IV = II − Ia		593	−382	671	882	671
						−139.4
V = III − Ib		889	1143	380	2412	
IVa			−573	1006	1322	531.6 $\zeta_2 = -0.896$ (Compression)
VI = V − IVa			1716	−626	1090	$\zeta_3 = +0.365$ (Tension).

Instead of calculating the values of ζ_2 and ζ_1 by substitution, as is done in the final column, it is better to repeat the calculation with a rearrangement of the columns. We interchange the first and third columns, and write the equations in reverse order for a reason still to be explained:

	ζ_3	ζ_2	ζ_1		s	
III	856	667	545	684	2753	
II	−286	667	−182	570	769	
IIIa		−223	−182	−228	− 919	
I	−429	−333	818	456	512	
IIIb		−333	−272	−342	−1376	
IV		890	0	798	1688 1888	$\zeta_2 = -0.897$
V		0	1090	798		$\zeta_1 = -0.732$

In this case ζ_2 is obtained immediately from IV. The small differences are reduced to the inaccuracy of slide rule calculations. Naturally, these will be so much smaller the smaller are the results of the multiplication, i.e., the quantities to be subtracted, since the relative errors (with a slide rule) are the same (3.11). It is advisable (and this is the basis of the above interchanges) to use that equation for the transformation and subtraction in which the unknown to be eliminated has the largest coefficient.

The same scheme is used for operations with a calculating machine, in which the accuracy is limited only by the number of digits on the machine. The calculation can also be carried out in this manner with logarithms. In this case the Gaussian logarithms are used for addition and subtraction purposes (4.7 and Art. 19).

3. It is of importance, especially for the method of least squares, that the following problem can be solved by the above scheme: we are to calculate *the value U of a linear form of n variables, which this form takes on for the n roots of a system of n linear equations.* We shall write out the scheme for only two unknowns, since the scheme for an arbitrary number of unknowns is entirely similar. The value of U is to be calculated from

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + a_{13} &= 0, \\ a_{21}x_1 + a_{22}x_2 + a_{23} &= 0, \\ a_{31}x_1 + a_{32}x_2 + a_{33} &= U, \end{aligned} \tag{8}$$

We now introduce another check. In addition to the sums s of the rows, we form the sums σ of the columns, and set $\sum s = \sum \sigma = \sum$. The scheme is then

	x_1	x_2		s
I	a_{11}	a_{12}	a_{13}	s_1
II	a_{21}	a_{22}	a_{23}	s_2
Ia		$\frac{a_{21}}{a_{11}} a_{12}$	$\frac{a_{21}}{a_{11}} a_{13}$	$\frac{a_{21}}{a_{11}} s_1$
III	a_{31}	a_{32}	a_{33}	s_3
Ib		$\frac{a_{31}}{a_{11}} a_{12}$	$\frac{a_{31}}{a_{11}} a_{13}$	$\frac{a_{31}}{a_{11}} s_1$
IV	σ_1	σ_2	σ_3	Σ
Ic		$\frac{\sigma_1}{a_{11}} a_{12}$	$\frac{\sigma_1}{a_{11}} a_{13}$	$\frac{\sigma_1}{a_{11}} s_1$
V = II − Ia		a'_{22}	a'_{23}	s'_2
VI = III − Ib		a'_{32}	a'_{33}	s'_3
Va			$\frac{a'_{32}}{a'_{22}} a'_{23}$	$\frac{a'_{32}}{a'_{22}} s'_2$
VII = IV − Ic		σ'_2	σ'_3	Σ'
Vb			$\frac{\sigma'_2}{a'_{22}} a'_{23}$	$\frac{\sigma'_2}{a'_{22}} s'_2$
VIII = VI − Va			a''_{33}	s''_3
IX = VII − Vb			σ''_3	Σ''

Here each σ is the sum of the coefficients directly above it; for example,

$$\sigma'_2 = \sigma_2 - \frac{\sigma_1}{a_{11}} a_{12} = \left(a_{12} - \frac{a_{11}}{a_{11}} a_{12}\right) + \left(a_{22} - \frac{a_{21}}{a_{11}} a_{12}\right)$$

(9)

$$+\left(a_{32} - \frac{a_{31}}{a_{11}} a_{12}\right) = a'_{22} + a'_{32} \, .$$

From this it follows as a final check that

$$a''_{33} = \sigma''_3 = s''_3 = \sum{}'',$$

and these values are equal to the desired value U if the correct values for x_1 and x_2 are substituted. Because the left side of the equation III has the value U, while that of the equation I_b has the value zero, subtraction of the latter does not change the value of U. The value of the left side of VI is then U, and since the value of Va is zero, the left side of VIII must also be U, upon subtraction of Va:

$$U = a''_{33} = \sigma''_3 = s''_3 = \Sigma''. \tag{10}$$

In many cases we deal with the co-called *symmetric equations,* in which the coefficients symmetric to the main diagonal have the same value. Then

$$a_{12} = a_{21} \, ; \quad a_{13} = a_{31} \, ; \quad a_{23} = a_{32} \, , \text{etc.}; \quad s_1 = \sigma_1 \, , \quad s_2 = \sigma_2 \tag{11}$$

etc. In this case, the equations which are obtained by elimination are also symmetric; for example,

$$a'_{32} = a_{32} - \frac{a_{31}}{a_{11}} a_{12} = a_{23} - \frac{a_{13}}{a_{11}} a_{21} - a'_{23} \, . \tag{12}$$

We can then omit the parts of the above scheme lying to the left, under the dotted line. The pattern of the equations used is indicated by the arrows.

These methods find their most important *application* in the *calculation of approximating curves.* Some discussions of the subject are found in Art. 25 and Art. 27, where an example is worked out. For a complete investigation of this field, the reader is referred to the extensive literature.[2] The calculations above find another application in the rather convenient *calculation of determinants.*

4. *How great are the errors, in the most unfavorable case, of the solutions* $x_1 \cdots x_n$, *if the coefficients and constants of the equations possess various errors?* If the coefficients a_{lm} have the errors $-\Delta_{lm}$, the constants a_l the errors $-\Delta_l$, then the unknowns x_l will have the errors $-\epsilon_l$. While the inexact values are determined from the equations

$$
\begin{aligned}
&a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n + a_1 = 0,\\
&a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n + a_2 = 0,\\
&\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\\
&a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n + a_n = 0,
\end{aligned} \tag{13}
$$

the correct values must satisfy the equations

$$
\begin{aligned}
&(a_{11} + \Delta_{11})(x_1 + \epsilon_1) + (a_{12} + \Delta_{12})(x_2 + \epsilon_2) + \cdots\\
&\qquad\qquad + (a_{1n} + \Delta_{1n})(x_n + \epsilon_n) + a_1 + \Delta_1 = 0,\\
&(a_{21} + \Delta_{21})(x_1 + \epsilon_1) + (a_{22} + \Delta_{22})(x_2 + \epsilon_2) + \cdots\\
&\qquad\qquad + (a_{2n} + \Delta_{2n})(x_n + \epsilon_n) + a_2 + \Delta_2 = 0,\\
&\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\\
&(a_{n1} + \Delta_{n1})(x_1 + \epsilon_1) + (a_{n2} + \Delta_{n2})(x_2 + \epsilon_2) + \cdots\\
&\qquad\qquad + (a_{nn} + \Delta_{nn})(x_n + \epsilon_n) + a_n + \Delta_n = 0.
\end{aligned} \tag{14}
$$

If we neglect the products $\Delta \cdot \epsilon$ and consider the equations (13), then we can estimate the errors of the unknowns from the equations

$$
\begin{aligned}
&a_{11}\epsilon_1 + a_{12}\epsilon_2 + \cdots\\
&\qquad\qquad + a_{1n}\epsilon_n + [\Delta_1 + \Delta_{11}x_1 + \Delta_{12}x_2 + \cdots + \Delta_{1n}x_n] = 0,\\
&a_{21}\epsilon_1 + a_{22}\epsilon_2 + \cdots\\
&\qquad\qquad + a_{2n}\epsilon_n + [\Delta_2 + \Delta_{21}x_1 + \Delta_{22}x_2 + \cdots + \Delta_{2n}x_n] = 0,\\
&\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\\
&a_{n1}\epsilon_1 + a_{n2}\epsilon_2 + \cdots\\
&\qquad\qquad + a_{nn}\epsilon_n + [\Delta_n + \Delta_{n1}x_1 + \Delta_{n2}x_2 + \cdots + \Delta_{nn}x_n] = 0.
\end{aligned} \tag{15}
$$

This is the same scheme which was used for the calculation of the x, except that the constants are different, and that, in getting upper limits for the errors, we add instead of subtracting in the case of the constants, while for the coefficients we take the absolute value and always subtract.

5. *Example*: We take the equations from the example in Sec. 2, in which the errors of the constants and the coefficients may be no larger than one half unit in the last place. Then we so choose the signs of the errors that the bracketed terms of the above equations reach their maximum values. The scheme then becomes

ϵ_1	ϵ_2	ϵ_3	
818	333	429	1.496
182	667	286	1.496
	74	96	0.333
545	667	857	1.496
	222	286	0.998
	593	190	1.829
	445	571	2.494
		143	1.373
		428	3.867

$\epsilon_3 \leq 0.0090.$

In the same way we calculate upper limits for the errors of the other unknowns:

857	667	545	1.496	
286	667	182	1.496	
	222	182	0.500	
429	333	818	1.496	
	333	273	0.750	
	445	0	1.996	$\epsilon_2 \leq 0.0045$
	0	545	2.246.	$\epsilon_1 \leq 0.0041.$

The percentage errors can then be no greater than

$$0.56\% \text{ for } \zeta_1\,, \qquad 0.5\% \text{ for } \zeta_2\,, \qquad 2.5\% \text{ for } \zeta_3\,.$$ [2]

A general formula for the upper limit of the error of the unknown, due to errors in the coefficients, has been given by Blumenthal.[3] For numerical calculations, the limits obtained above are usually too large, but the estimates can be of use in purely theoretical considerations, as the given examples show.

6. In technical work, linear equations occur particularly in *structural mechanics*. For example, in the investigation of a framework or of a cross-beam for various loads, we frequently have the case that the coefficients of the unknowns determined by the structure remain unchanged, while we must substitute various values for the constant term, as determined by the loads. These latter terms are therefore designated by load numbers,

If we have a large number of such groups of equations, then the following procedure is advisable. We set all the constants equal to zero except for a_1, to which is given the value 1. The values of x obtained are denoted by $x_\kappa^{(1)}$ where κ has the values 1 to n. Then we set all the constants except a_2 equal to zero, and give a_2 the value 1. The solutions of the system of equations are then designated by $x_\kappa^{(2)}$, etc. The calculation is always carried out by means of the following scheme, in which only the last column changes each time. If a_r has the value a_r instead of 1, then the solutions would be a_r times as great. Since we are dealing with linear equations, these solutions could be superimposed. The general solution for the system of equations with constant terms $a_1 \cdots a_n$ would then be

$$x_\kappa = \sum_{r=1}^{n} a_r x_\kappa^{(r)}. \tag{16}$$

7. *The determination of the unknowns* from linear equations *can also be carried out graphically*[4]. In this, construction of expressions of the form

$$y = a_l + a_{l1}x_1 + a_{l2}x_2 + \cdots + a_{ln}x_n \tag{17}$$

is necessary. We first construct a graph for such an expression. In this graph we place the lengths a_l, a_{l1}, a_{l2}, $\cdots$, a_{ln} on a straight line, perhaps the x axis, beginning from a point O, and measured in the direction determined by their signs. The starting point of each length is set at the

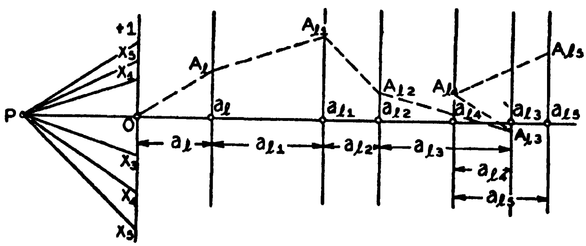


FIG. 91

endpoint of the previous one, as is shown in Fig. 91. Here only a_{l4} is negative. We draw parallels to the y axis through the endpoints of these lengths. On the y axis we draw, from the origin, the lengths $1, x_1, x_2, \cdots$, for which we want to construct the value of the above linear form. It is not necessary that the scale modulus used here be the same as the one used for the x axis. Lines are drawn connecting these points with the point -1 on the x axis, the pole P. To the individual rays of this pencil of guiding lines, we draw parallels, starting from the origin, in each corresponding section of the framework of the linear form. Then $OA_l \parallel P_1$; then $a_lA_l = a_l$, measured in the units of the y axis. Furthermore, $A_lA_{l1} \parallel Px_1$, and therefore $a_{l1}A_{l1} = a_l + a_{l1}x_1$, etc. The length

$$a_{ii}A_{ii} = a_i + a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{ii}x_i$$

is cut off on the $(i + 1)$st y-parallel. If $y = 0$, i.e., if the x_m have values which satisfy the equation

$$a_i + a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n = 0, \tag{18}$$

then the endpoint of the polygon with the connecting points A must lie on the x axis, indeed at a_{in}. If the x_m should therefore be solutions of n equations with n unknowns, we have the following result. The parallels to the corresponding straight lines of the pencil of rays must give a polygon, in each of the n schemes of the n equations, the endpoint A_{in} of which falls on the x axis at the point a_{in} of the scheme concerned.

It is easy to accomplish this for a *system of reduced equations,* such as we get by the elimination of a system of ordinary equations. For example, we would have for three unknowns,

$$\begin{aligned} 0 &= a_1 + a_{11}x_1 , \\ 0 &= a_2 + a_{21}x_1 + a_{22}x_2 , \\ 0 &= a_3 + a_{31}x_1 + a_{32}x_2 + a_{33}x_3 . \end{aligned} \tag{19}$$

Here the direction of x_1 can be found from the scheme of the first equation,

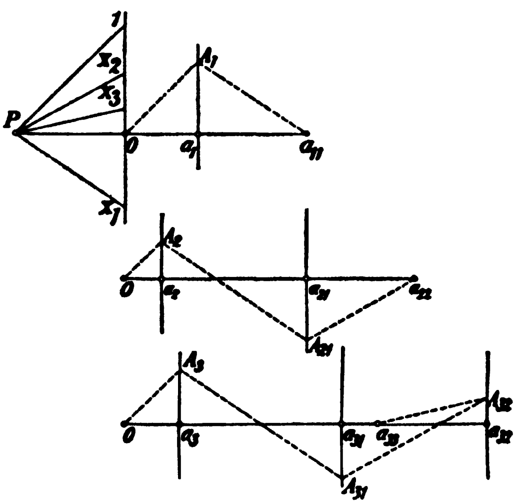


FIG. 92

the direction of x_2 from the second scheme, etc., as can be seen from Fig. 92, in which the three equations

$$0 + 2 + 3x_1 \, ,$$

(19a) $$0 = 1 + 4x_1 + 3x_2 \, ,$$

$$0 = 1.5 + 4.5x_1 + 4x_2 - 3x_3$$

are represented. It is therefore possible, by graphical elimination, to transfer an arbitrary system of equations into a system of related equations. The unknowns can easily be found from these. From the drawing we read off

$$x_1 = -0.667; \qquad x_2 = 0.556; \qquad x_3 = 0.240.$$

8. Van den Berg[5] has given *a graphical elimination method.* This permits the construction of the $n - 1$ frameworks of $n - 1$ equations with $n - 1$ unknowns from n frameworks with n equations for n unknowns. If we have, for example, the schemes of the two equations

(20) $$\begin{aligned} a_l + a_{l1}x_1 + a_{l2}x_2 + \cdots + a_{ln}x_n &= 0 \\ a_m + a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= 0 \end{aligned}$$

plotted on lines parallel to the x axis, if the points a_{li} and a_{mi} with the

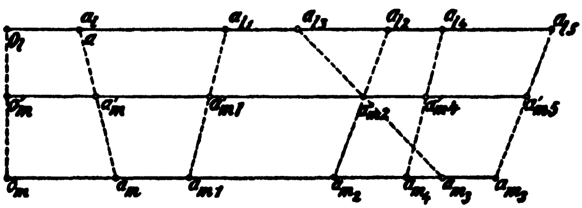


FIG. 93

corresponding indices i are connected together, and finally, if an x-parallel is drawn, on which the connecting lines cut segments of length

(20a) $$a'_m \, , \, a'_{m1} \, , \, a'_{m2} \, , \, \cdots \, , \, a'_{mn} \, ,$$

then we can calculate the lengths of these segments from the proportion

(20b) $$(a_{mi} - a'_{mi}) : (a'_{mi} - a_{li}) = 0_m0'_m : 0'_m0_l = \lambda.$$

This gives

(21) $$a'_{mi} = \frac{a_{mi} + \lambda a_{li}}{1 + \lambda}.$$

If we now draw the x-parallel through the intersection of two connecting lines, then the values of a'_{mi} are zero, i.e., we have $\lambda = -a_{mi}/a_{li}$; therefore the segments cut off on this x parallel become

$$a_{m\kappa} - \frac{a_{mi}}{a_{li}} \cdot a_{l\kappa} = a'_{m\kappa} , \tag{22}$$

except for a constant factor. But these are the coefficients which are obtained from the first elimination in the Gaussian scheme. The coefficients of x_κ become zero if we draw the parallels all through the corresponding points of intersection. Therefore this unknown is eliminated. From these $n - 1$ equation frameworks, we can produce $n - 2$ schemes for $n - 2$ unknowns, etc. If we then choose one scheme from each group, then we have a system of n related equations, from which we can determine the unknowns by the method explained in the preceding section.

In order to obtain satisfactory intercepts in the elimination process, we observe that each of the equations can be multiplied by an arbitrary positive or negative number, that the origin can be arbitrarily chosen, and finally, that the distance between the x-parallels can be chosen appropriately.

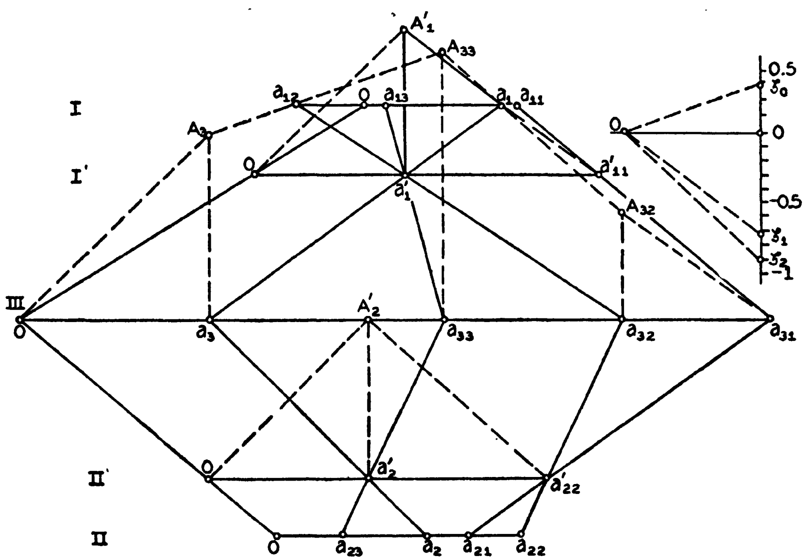


FIG. 94

9. *Example*: As an example, we consider (Fig. 94) the three equations set forth in Sec. 2. With a suitable choice of the origin, the schemes of the three equations

$$\text{I.}\quad 456 - 429\zeta_3 - 333\zeta_2 + 818\zeta_1\,,$$

$$\text{(22a)}\qquad \text{II.}\quad 570 - 286\zeta_3 + 667\zeta_2 - 182\zeta_1\,,$$

$$\text{III.}\quad 684 + 857\zeta_3 + 667\zeta_2 + 545\zeta_1$$

are drawn, one under the other, at equal distances. Corresponding points of I-III and II-III are connected, and the x-parallels I′ and II′ are so placed that the coefficients of x_3 become zero. For this reason, the x-parallel II′ also goes through the intersection of the lines $a_{21}a_{31}$ and $a_{22}a_{32}$ so that the coefficient of x_1 in this equation is also zero, and so that a_1a_3, $a_{13}a_{33}$ and $a_{12}a_{32}$ go through one point, so that in the equation whose scheme is I′, the coefficients of x_2 and x_3 are zero. It is a special advantage of the graphical method that such a calculation as the above is easier than the calculation in the numerical solution of the equations. We find $\zeta_2 = -0.896$ in the adjacent figure by means of the dotted line $OA_2'a_{22}'$ above II′; through the polygon $OA_1'a_{11}'$ over I′, we find $\zeta_1 = -0.730$. Finally, we draw the polygon sides OA_3 and $a_{31}A_{32}A_{33}$ in the frame of the equation III, by use of the directions which have been found. The final line A_3A_{33} gives the direction for $\zeta_3 = 0.365$. The values agree very well with those calculated in Sec. 2. If the accuracy of the values found by the drawing is not sufficient, we can produce new equations, suitable for the corrections, just as is done in the error considerations in Sec. 6. These equations have exactly the same schemes, except for the now very small values a_l; we can then use the same scheme for the correction calculation, except that we measure the value a_l in a scale which is 100 times greater. The corrections which are then found are to be divided by one hundred.

10. *Another method for the construction of the solutions of n linear equations with n unknowns from two solutions of n* − 1 *equations with n unknowns has been given by Massau.*[6] For his construction, we need the following theorem:

If we have any system $x_1, x_2, \cdots, x_n$ of solutions (of which there are an infinite number) of $n - 1$ equations

$$\begin{aligned}
&a_1 + a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = 0,\\
&a_2 + a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = 0.\\
&\quad \cdots\cdots\cdots\cdots\cdots\cdots\\
&a_{n-1} + a_{n-1,1}x_1 + a_{n-1,2}x_2 + \cdots + a_{n-1,n}x_n = 0,
\end{aligned}\tag{23}$$

and construct two linear forms of these n quantities

$$(24)\qquad \begin{aligned} X &= b_1x_1 + b_2x_2 + \cdots + b_nx_n \,, \\ Y &= c_1x_1 + c_2x_2 + \cdots + c_nx_n \,, \end{aligned}$$

then there exists the linear relation among these quantities

$$(25)\qquad lX + mY + n = 0.$$

This is seen immediately, if we observe that the determinant

$$(26)\qquad \Delta = \begin{vmatrix} -X & b_1 & b_2 & \cdots & b_n \\ -Y & c_1 & c_2 & \cdots & c_n \\ a_1 & a_{11} & a_{12} & \cdots & a_{1n} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{n-1} & a_{n-1,1} & a_{n-1,2} & \cdots & a_{n-1,n} \end{vmatrix} = 0$$

must be zero in order that the $n + 1$ equations (23) and (24) be satisfied for the unknowns. Besides, it is seen that the quantities l and m are independent of the constants $a_1 , \cdots , a_{n-1}$. If we now take the scheme of the nth equation

$$(27)\qquad a_n + a_{n,1}x_1 + a_{n,2}x_2 + \cdots + a_{n,n}x_n = 0$$

and construct on this the polygon with n directions x_m , which satisfy the $n - 1$ equations (23), then the $(i - 1)$st and the ith junction points have the ordinates

$$(28)\qquad \begin{aligned} y_{i-1} &= a_n + a_{n,1}x_1 + a_{n,2}x_2 + \cdots + a_{n,i-1}x_{i-1} \,. \\ y_i &= a_n + a_{n,1}x_1 + a_{n,2}x_2 + \cdots + a_{n,i-1}x_{i-1} + a_{ni}x_i \,. \end{aligned}$$

If we now set $X = y_{i-1} - a_n$; $Y = y_i - a_n$, then, by the above theorem, the linear equation

$$(29)\qquad ly_{i-1} + my_i + \bar{n} = 0$$

exists, where again l and m are independent of the constant terms a_1 , $a_2 , \cdots , a_n$. This relation holds for each arbitrary value i from 1 to n; but then also for all the infinitely many sets of value $x_1 , x_2 , \cdots , x_n$ which satisfy the equations (23). Among these, there could be one for which $y_{i-1} = 0$; then by (29), $y_i = a_iC = -\bar{n}/m = \mu$; another for which $y_i = 0$, so that $y_{i-1} = a_{i-1}B = -\bar{n}/l = \nu$. The two lines $a_{i-1}C$ and Ba_i intersect at the point P_i , and it can be shown that for each line through P_i the

equation (29) is satisfied. From Fig. 95, for example, we read off for the straight line $A_{i-1}A_i$,

(30) $$y_{i-1} : \nu = (\mu - y_i) : \mu$$

or $\mu y_{i-1} + \nu y_i - \mu\nu = 0$. If we substitute the values of μ and ν, then the equation (29) is obtained. Therefore we denote this equation as the equa-

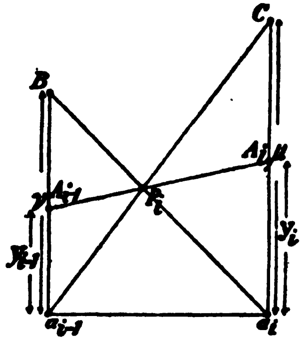


FIG. 95

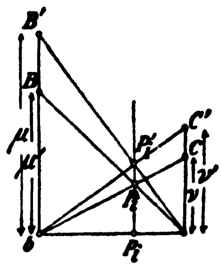


FIG. 96

tion of the point P_i in parallel coordinates. From Fig. 96, in which μ and μ' are interchanged, we derive the fact that for points on a y-parallel,

(30a) $$\frac{\mu}{\mu'} = \frac{p_iP_i}{p_iP'_i} = \frac{\nu}{\nu'},$$

i.e., if we substitute the values,

(30b) $$\frac{m}{l} = \frac{m'}{l'}.$$

Therefore, if the ratio of m/l, in the equations of the various points, is constant, then these points lie on a parallel to the y axis.

But the equation (29) assumes that the direction of the line $A_{i-1}A_i$ corresponds to a system of solutions of the equations (23). Therefore, if we have any system of solutions of the equations (23), and if we draw the corresponding polygon in the scheme of the equation (27), then all polygon sides go through the junction point P_i. Now, since i can take on the values from 1 to n, to each interval there is one such a junction point through which pass all the polygon sides of the interval, if the $x_1, \cdots, x_n$ are one of the infinitely many systems of solutions of (23).

If the determinant of the coefficients of the n equations (23) and (27), with the n unknowns, $x_1, x_2, \cdots, x_n$, is not zero, then there is one and only one set of solutions $x_1, \cdots, x_n$ of these n equations. Therefore, among the infinitely many polygons which can be drawn in the scheme

of the nth equation, and which satisfy the equations (23), i.e., go through the junction points $P_2, \cdots, P_n$, there must be one for which $y_n = 0$, for which, therefore, A_{nn} and a_{nn} coincide. But this polygon is uniquely determined by the points $a_{nn}, P_n, \cdots, P_2, A_n, 0$ (Fig. 97).

If we then have two solutions of the equations (23), by means of which the junction points $P_2, \cdots, P_n$ are determined in the scheme of the equation (27), then the polygon drawn out from a_{nn} through these junction

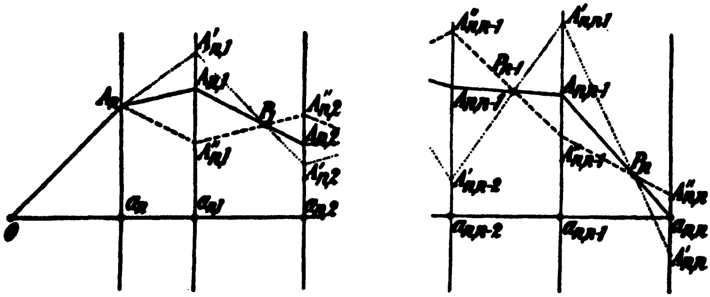


FIG. 97

points gives the system of solutions of the n equations (23) and (27). Then the solution of n equations with n unknown is reduced to the investigation of two solutions of a system of $n - 1$ equations with $n - 1$ unknowns, which we obtain by assuming an arbitrary value for one of the unknowns in the equations (23).

11. *The method becomes considerably simpler* if we can produce from n solutions of the first equation (if these satisfy certain conditions) $n - 1$ solutions of the first two, and from these, $n - 2$ solutions of the first three equations, etc.

In this case we must observe the following. If we have two sets of solutions L'_{m-1} and L''_{m-1} of the first $m - 1$ equations which coincide in $n - m$ directions, then a set of solutions L'_m of the first m equations can be constructed from these. In these solutions, the first $n - m$ directions remain unchanged. If the first $n - m$ directions determine the lengths

$$\bar{a}_1 = a_{1,n-m}A_{1,n-m}; \quad \bar{a}_2 = a_{2,n-m}A_{2,n-m}; \quad \cdots ; \quad \bar{a}_m = a_{m,n-m}A_{m,n-m},$$

on the $(n - m + 1)$st y-parallels of the first m equations, then we can regard these quantities as the constants of m equations whose mth scheme is the last part of the framework of the first m equations. These equations become

$$(31) \qquad \bar{a}_1 + a_{1,n-m+1}x_{m+1} + a_{1,n-m+2}x_{m+2} + \cdots + a_{1,n}x_n ,$$

$$\bar{a}_2 + a_{2,n-m+1}x_{m+1} + a_{2,n-m+2}x_{m+2} + \cdots + a_{2,n}x_n ,$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\bar{a}_m + a_{m,n-m+1}x_{m+1} + a_{m,n-m+2}x_{m+2} + \cdots + a_{m,n}x_n .$$

For the first $m - 1$ of these equations, the last m directions of L'_{m-1} and L''_{m-1} give two sets of solutions, from which we can find one solution L'_m for all m equations (31), and thereby, if we add the first $n - m$ common directions of L'_{m-1} and L''_{m-1}, for the first m equations of the systems (23) and (27).

We further observe that the junction points P_i are shifted on a parallel to the y axis if the ratio in the equation (29) remains unchanged. Also it was shown in Sec. 10 that l and m do not depend on the constants of the equations, and therefore, applied to the system (31), do not depend on $\bar{a}_1 , \cdots , \bar{a}_m$. Then we see that the junction points (at which points $A'_{1,n-m} , A'_{2,n-m} , \cdots , A'_{m,n-m}$, the $(n - m + 1)$st y-parallel is always cut by the polygon) of the last $m - 1$ intervals, for all sets of solutions of the first $m - 1$ equations, must lie on the y-parallel through the junction points which are determined by the intersection of the corresponding sides of the sets of solutions L'_{m-1} and L''_{m-1}. Each additional set of solutions L'''_{m-1} of the first $m - 1$ equations, introduced in the scheme of the mth equation, then yields a new set of solutions L''_m of the first m equations. This set coincides in the directions of the first $n - m$ intervals with those of L'''_{m-1}. In the last $m - 1$ intervals, the intersections of the polygon sides of L'''_{m-1}, with the y-parallels through the junction points give a new set of junction points through which we can draw a new solution L''_m of the first m equations. We draw a polygon through these new junction points $P'_n , P'_{n-1} , \cdots , P'_{n-m+2}$ beginning at a_{mn} and ending at $A'_{m,n-m}$.

From r polygons which satisfy the first $m - 1$ equations, and two of which coincide in the first $n - m$ directions, $r - 1$ polygons can be constructed, which satisfy the first m equations, and which have the same directions, in the first $n - m$ intervals, as the r polygons from which we began.

We must then take care that from n systems of solutions of the first equation, of which two coincide in $n - 2$ directions, we get $n - 1$ sets of solutions of the first two equations, of which two coincide in the first $n - 3$ directions. But we get these whenever at least three sets of solutions of the first equation coincide in the first $n - 3$ directions. We get the two solutions of the first three equations, which have the first $n - 5$ directions in common from four solutions of the first equation, whose first $n - 4$ directions coincide, etc. For five equations, for example, we could start

from the five polygons given in Fig. 98. In this figure coincident points are only drawn once, and the lengths are omitted.

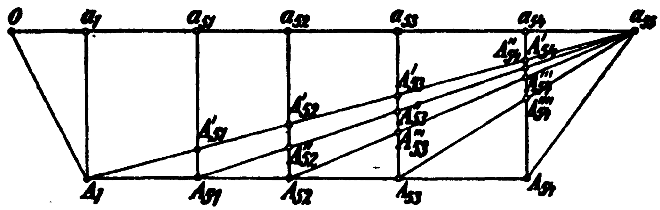


FIG. 98

In certain forms of the equations, especially for difference equations (cf. Art. 24), the process becomes especially simple, as an example shows.

12. *Example*: A transverse beam rests on 6 supports which have the distances $l_1 = 3m$, $l_2 \doteq 4m$, $l_3 = 5m$, $l_4 = 3m$, $l_5 = 4m$ between successive supports. In the first, fourth and fifth sections, the beam is loaded with 50 kg., in the second and third sections with 500 kg. The support moments are to be calculated.

The bending moments about the supports, the so-called support moments, are zero at the end points 0 and 5. They can be calculated at the intermediate points 1, 2, 3, and 4 from the Clapeyron equations.[7] In our case, we get the equations

$$\begin{aligned}
14M_1 + 4M_2 \qquad\qquad\qquad &= -\ 8337\ 1/2 \\
4M_1 + 18M_2 + 5M_3 \qquad\quad &= -23625 \\
5M_2 + 16M_3 + 3M_4 &= -15962\ 1/2 \\
3M_3 + 14M_4 &= -\ 1137\ 1/2.
\end{aligned}$$

Because of the steplike structure of these equations, we get two solutions, one of which is a dashed line in Fig. 99, the other dotted. The final solution is then obtained. The portion for the constants is omitted; this can be similarly plotted on the first y-parallel in corresponding scale. From the figure at the top, we read off the solutions

$$M_1 = -290, \qquad M_2 = -1050, \qquad M_3 = -670, \qquad M_4 = +60.$$

If we substitute these values as first approximations in the equations above, then we get new equations for the calculation of the corrections. We can consider these by use of the same scheme just as in the first case. In this case we naturally choose a corresponding scale for

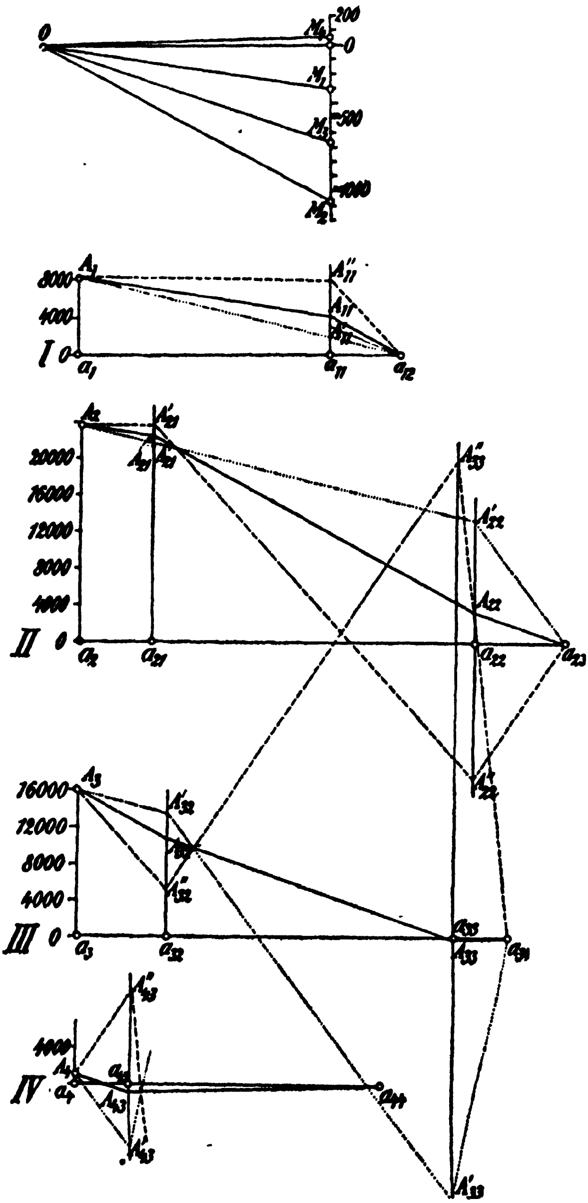


FIG. 99

the constants. The shearing forces and the reactions of the supports can be calculated from the support moments.

Equation instruments have also been constructed for the solution of linear equations with several unknowns, for example those by Thomson.[8]

NOTES

1. Pasternack, *Berechnung vielfach statisch unbestimmter biegefester Stab- und Flachentragwerke* (Zürich, 1926).
2. For example, Helmert, *Die Ausgleichungsrechnung.* 2nd ed. (1924); Czuber, *Wahrscheinlichkeitsrechnung und ihre Anwedung auf Fehlerausgleichung*, I. 4th ed. (1924); Happach, *Ausgleichungsrechnung* (Leipzig, 1923); Weitbrecht, *Ausgleichungsrechnung*, I and II (Berlin, 1919-20).
3. Blumenthal, *Z. f. Math. u. Phys.* 62 (1914), p. 360.
4. A method not described here is given by Mehmke, *Leitfaden zum graphischen Rechnen* (Berlin, 1917).
5. Van den Berg. *Verl. med. v. d. Kon. Acad. v. Wet.* 3rd series, 6 (1888).
6. Massau, *Ann. de l'Assoc. des ing. sortis des écoles spéciales de Gand* (1889).
7. Cf. textbooks of mechanics and graphical statics, e. g., those by Müller-Breslau, Föppl, Lorenz, etc.
8. Jacob, *Calcul mécanique* (Paris, 1911); Herzog-Feldmann, *Die Berechnung elektrischer Leitungsnetze* (1927).

24. Linear Difference Equations.

1. *Difference equations*, which have the form

$$f_r(y_r\ ,\ \Delta^1_{r+(1/2)}\ ,\ \Delta^2_{r+1}\ ,\ \cdots\ ,\ \Delta^m_{r+(m/2)}) = 0\ (r = 0, 1, \cdots, n) \tag{1}$$

are a special type of the equations with several unknowns; the Δ^r here have the meaning given in 10.2. If we substitute the values for these given in 10(1) to 10(4), then we get from (1)

$$f_r(y_r\ ,\ y_{r+1}\ ,\ \cdots\ ,\ y_{r+m}) = 0\ (r = 0, 1, \cdots, n). \tag{2}$$

The values y_r , y_{r+1} , $\cdots$ can be interpreted as values of a function for equidistant arguments x_r , x_{r+1} , $\cdots$. In the following therefore, we usually speak of function values.

If y_1 and y_{r+m} actually appear in the equation thus formed, we characterize the equations as difference equations of the mth order. In general, n is larger than m. In each successive equation of the system, the function value y_r which had the lowest index in the previous equation, is lacking, and there appears a new function value y_{r+m+1} with an index which is one unit higher than the largest index of the preceding equation. Therefore $n + m + 1$ unknowns appear in $n + 1$ equations. Of these, m are arbitrarily chosen. The number of arbitrary constants is therefore equal to the degree of the equation. If these are the first m values y_0 , $\cdots$, y_{m-1}

or the last m values $y_{n+1}, \cdots, y_{n+m}$, then we can calculate the other unknowns, step by step. This is done in the first case from y_m, and in the second from y_{m+1}, etc. We can therefore speak of a *recursion formula* in this case. A particular solution in which

$$y_0^{(k)} = c_0, y_1^{(k)} = c_1, \cdots, y_{m-1}^{(k)} = c_{m-1} \tag{2a}$$

are given arbitrarily, is designated by

$$y_r^{(k)}|_{c_0, c_1, \cdots, c_{m-1}} . \tag{2b}$$

Such a particular solution in which we start from the end values $y_{n+1}^{(k)} = c_{n+1}, \cdots, y_{n+m}^{(k)} = c_{n+m}$ is designated by $y_r^{(k)} \mid^{c_{n+1}, \cdots, c_{n+m}}$. In this case, r has all values from 0 to $n + m$.

2. In the difference equations appearing in the applications, especially in technical work, the given function values y_r do not lie only at the beginning, but at the beginning and the end. Therefore the y_r with highest and lowest indices are given. This case occurs in *boundary value problems*. In practice, one usually deals with linear difference equations which can be written in the form

$$\sum_{l=q_1}^{l=q_2} a_{r,l+r} y_{l+r} = a_r \qquad (r = 1, 2 \cdots n), \tag{3}$$

where $q_1 = -(m + 1)/2$, $q_2 = (m - 1)/2$ for odd m, and $q_1 = -m/2$, $q_2 = +m/2$ for even m. If the constants a_r on the right side are not all zero, the equations are called *inhomogeneous linear difference equations*; if these constants are all zero, we have the corresponding homogeneous linear difference equations.

A series of theorems exist for these linear difference equations which correspond to those of linear differential equations. If, for example, we have two particular solutions of the equation (3), $y_x^{(l)}$ and $y_x^{(k)}$, then, as may be seen by substitution and subtraction, $\eta_x = y_x^{(l)} - y_x^{(k)}$ is a solution of the corresponding homogeneous system of equations.

The *superposition principle*, formulated in 23.6, is also valid for the linear difference equations. If in particular, we have homogeneous equations, then each linear combination of the solutions is also a solution of the system of equations. If we have s solutions $\eta_x^{(r)} (r = 0, 1, \cdots, s - 1)$ of the homogeneous equations

$$\sum_{l=q_1}^{l=q_2} a_{r,l+r} \eta_{l+r} = 0 \qquad (r = 1, 2, \cdots, n), \tag{4}$$

then these s solutions are said to be linearly dependent if for all values of x there exists among them the relation

(5)
$$\sum_{r=0}^{r=s-1} A_r \eta_x^{(r)} = 0$$

in which not all the A_r are zero. If there is no such system of constants A_r, then the solutions are linearly independent. The m linearly independent solutions of a homogeneous linear difference equation of mth order form a *complete set of solutions*, because every other solution can be represented as a linear combination of these solutions:

(6)
$$\eta_x = \sum_{r=1}^{r=m} A_r \eta_x^{(r)}.$$

If the m solutions are independent, then for m successive values of η_x, perhaps for $x = p, \cdots, p + m - 1$ the A_r could so be determined that these m values of η_x are 0. But if m consecutive values of η_x are zero, as follows from the form of the linear homogeneous difference equations, all values of η_x are zero. But the equations

(6a)
$$\sum_{r=1}^{r=m} A_r \eta_x^{(r)} = 0$$

are consistent under these conditions only if

(7)
$$\Delta = \begin{vmatrix} \eta_p^{(1)} & \eta_{p+1}^{(1)} & \cdots & \eta_{p+m-1}^{(1)} \\ \eta_p^{(2)} & \eta_{p+1}^{(2)} & \cdots & \eta_{p+m-1}^{(2)} \\ \cdot & \cdot & \cdot & \cdot \\ \eta_p^{(m)} & \eta_{p+1}^{(m)} & \cdots & \eta_{p+m-1}^{(m)} \end{vmatrix} = 0.$$

If $\Delta \neq 0$, we have a system of independent equations. In this case the coefficients A_r can be calculated from (6) for arbitrary η_x, because $\Delta \neq 0$ is the condition that these m equations can be solved. Now, since each solution is determined by m successive values, then each solution of the homogeneous difference equations can be represented as a combination of the m solutions of the complete set.

The *solution of the inhomogeneous equation* can be reduced to an integration by means of a *complete set of solutions* of the corresponding homogeneous equation, according to the method of the *variation of constants*, given by Lagrange.[2] The reader is referred to textbooks of the calculus of differences.[3]

3. In practice we are generally concerned with boundary value problems, as has been mentioned previously. Suppose that the k constants $y_{q_1}, y_{q_1+1}; \cdots, y_{q_1+k-1}$ are given as initial conditions, and the other

constants $y_{a_0+k+1}, \cdots, y_{a_0+m}$ as final conditions. This problem can be solved for the inhomogeneous difference equation by use of solutions of the joint difference equation for m successive values. If we have the linear difference equations (3), and if the series is multiplied by $\xi_1, \xi_2, \cdots \xi_n$, added together, and arranged according to the factors of y_k, then these factors, except for the first and last, have $m + 1$ terms, and have the form

$$a_{r,l+r}\xi_r + a_{r+1,l+r}\xi_{r+1} \cdots a_{r+m,l+r}\xi_{r+m}\,. \tag{8}$$

If the resultant expression is equated to zero, then these are the homogeneous difference equations adjoint to the original equation.

4. We shall illustrate the application of the adjoint difference equation to the *inhomogeneous linear difference equation of second order* which, for example, plays a role in calculations on a transverse support. As boundary conditions, let $x_0 = x_{n+1} = 0$; then we have the equations

$$\begin{array}{lr} a_{11}x_1 + a_{12}x_2 & = a_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 & = a_2 \\ \qquad a_{32}x_2 + a_{33}x_3 + a_{34}x_4 & = a_3 \\ \cdots\cdots\cdots\cdots\cdots\cdots & \\ a_{n-1,n-2}x_{n-2} + a_{n-1,n-1}x_{n-1} + a_{n-1,n}x_n & = a_{n-1} \\ + a_{n,n-1}\ x_{n-1} + a_{n,n}x_n & = a_n\,. \end{array} \tag{9}$$

The corresponding adjoint system is

$$\begin{array}{lr} a_{11}\xi_1 + a_{21}\xi_2 & = 0 \\ a_{12}\xi_1 + a_{22}\xi_2 + a_{32}\xi_3 & = 0 \\ \qquad a_{23}\xi_2 + a_{33}\xi_3 + a_{43}\xi_4 & = 0 \\ \cdots\cdots\cdots\cdots & \end{array} \tag{10}$$

By a stepwise procedure, we determine a solution $\xi_r|_{0,\alpha}$ of the equations (10) for the initial conditions $\xi_0 = 0, \xi_1 = \alpha$. We then multiply the equations (9) successively by $\xi_1, \cdots$ etc., and add. We obtain

$$x_n(a_{n-1,n}\xi_{n-1}|_{0,\alpha} + a_{n,n}\xi_n|_{0,\alpha}) = \sum_{r=1}^{r=n} a_r\xi_r|_{0,\alpha}\,. \tag{11}$$

From these, x_n is calculated, and the other unknowns are calculated by reversing the procedure step by step. Naturally this x_1 could just as well have been calculated first with a particular solution $\xi_r|^{0,\beta}$.

Other *boundary conditions*, for example, $x_0 + \lambda x_1 = k$ can be reduced to the above. We need only set this equation at the beginning of the system and consider it as a difference equation. If we then add $x_1 = 0$ as a boundary condition, we have reduced this case to the above. To be sure, this is not always possible. For example, this cannot be done with boundary conditions which are encountered in difference equations for frameworks which are connected cyclically.

Example: We again consider the example of the transverse support, already treated graphically in 23.12 by the method of Massau. This method is also suitable in the simplification given there for the handling of difference equations. In this example, coefficients which are symmetric with respect to the main diagonal are equal; i.e., $a_{i\kappa} = a_{\kappa i}$. Consequently the system of adjoint difference equations coincides with the homogeneous set belonging to the given equations; we say that these equations are self-adjoint. We have the equations

$$\begin{aligned}
14M_1 + 4M_2 \qquad\qquad\qquad &= -\ 8{,}337\ 1/2\\
4M_1 + 18M_2 + 5M_3 \qquad\quad &= -23{,}625\\
5M_2 + 16M_3 + 3M_4 &= -15{,}962\ 1/2\\
3M_3 + 14M_4 &= -\ 1{,}137\ 1/2.
\end{aligned}$$

We also add $M_0 = M_5 = 0$, so that we have a system of linear difference equations of second order which are self-adjoint. The system of solutions of the adjoint equations is

$$\xi_1 = \alpha, \qquad \xi_2 = -\frac{7}{2}\alpha, \qquad \xi_3 = +11.8\alpha, \qquad \xi_4 = -57.1\alpha.$$

From these we obtain, upon dividing by α,

$$M_4(35.4 - 799.4)$$

$$= (-8{,}337.5 + 82{,}687.5 - 188{,}357.5 + 64{,}951.25)$$

$$M_4 = \frac{49{,}056.25}{764} = 64.2098.$$

We calculate

$$M_3 = -678.8, \qquad M_2 = -1058.8, \qquad M_1 = -293.0,$$

stepwise. As proof, we substitute M_1 and M_2 in the first equation. If we use the values rounded off above, this gives

$$-8337.2 \approx -8{,}337.5,$$

which is in good agreement. If we do not round off, then both sides are identical.

The calculations for *difference equations of higher order* is the same as above, except that we must use several sets of solutions of the adjoint equations.

5. In 18.8 it was shown how a system of linear equations could be solved by iteration if the coefficients of the main diagonal are predominant. Occasionally, it happens that not only are there several other coefficients which are large, but that the coefficients in the two diagonals adjacent to the main diagonal are of the same order of magnitude as those of the terms in the main diagonal, and that all other coefficients are small in comparison. We can then omit the terms with small coefficients to begin with:

$$\begin{array}{llllllll}
a_{11}x_1 + a_{12}x_2 & + a_{13}x_3 + a_{14}x_4 \cdots & & & + a_{1n}x_n & = a_1 \\
a_{21}x_1 + a_{22}x_2 + a_{23}x_3 & + a_{24}x_4 \cdots & & & + a_{2n}x_n & = a_2 \\
a_{31}x_1 & + a_{32}x_2 + a_{33}x_3 + a_{34}x_4 & \cdots & & + a_{3n}x_n & = a_3 \\
\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot & & & & & \\
a_{n-1,1}x_1 \cdots & + a_{n-1,n-2}x_{n-2} + a_{n-1,n-1}x_{n-1} + a_{n-1,n}x_n & & & & = a_{n-1} \\
a_{n,1}x_1 \cdots & + a_{n,n-2}x_{n-2} & + a_{n,n-1}x_{n-1} & + a_{n,n}x_n & & = a_n .
\end{array} \tag{12}$$

We then consider the system of bracketed terms which remain as difference equations of second order. These can be solved by the methods given above. If we then write

$$\begin{aligned}
a_{11}x_1 + a_{12}x_2 \qquad\qquad &= a_1 - a_{13}x_3 - a_{14}x_4 \cdots - a_{1n}x_n \\
a_{21}x_1 + a_{22}x_2 + a_{23}x_3 \qquad &= a_2 - a_{24}x_4 - a_{25}x_5 \cdots - a_{2n}x_n \\
a_{32}x_2 + a_{33}x_3 + a_{34}x_4 &= a_3 - a_{31}x_1 - a_{35}x_5 \cdots - a_{3n}x_n \\
\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot &
\end{aligned} \tag{13}$$

and substitute the approximate values $x_1^{(1)} \cdots x_n^{(1)}$ (initially calculated) on the right side, we have a new system of difference equations in which only the right sides are changed. Therefore in equation (11), only the right side is to be calculated with the same $\xi_r|_{0,\alpha}$. We then get new approximation values $x_1^{(2)} \cdots x_n^{(2)}$, with which we deal in exactly the same manner. This procedure is continued until no further changes occur in the approximation values.[5]

Example: The equations

$$\begin{array}{lllllll}
12x_1 & + 5x_2 & + 0.3x_3 & + 0.8x_4 & + 0.4x_5 & + 0.5x_6 & = 10 \\
6x_1 & + 15x_2 & + 7x_3 & + 0.8x_4 & + 1x_5 & + 0.3x_6 & = 16 \\
0.9x_1 & + 5x_2 & + 16x_3 & + 8x_4 & + 0.1x_5 & + 0.6x_6 & = 18 \\
0.4x_1 & + 1.1x_2 & + 11x_3 & + 16x_4 & + 4x_5 & + 0.3x_6 & = 20 \\
0.2x_1 & + 0.7x_2 & + 1.2x_3 & + 7x_4 & + 14x_5 & + 5x_6 & = 17 \\
0.7x_1 & + 0.5x_2 & + 0.7x_3 & + 1.3x_4 & + 8x_5 & + 10x_6 & = 14
\end{array}$$

are given. The adjoint system of the bracketed difference equations has the solution

(13b $\quad \xi_1 = 1, \; \xi_2 = -2, \; \xi_3 = +5, \; \xi_4 = -6, \; \xi_5 = +8, \; \xi_6 = -11.$

With these we get a first approximation

$$\text{(13c)} \qquad \begin{array}{lll}
x_6^{(1)} = 1, & x_5^{(1)} = 0.5, & x_4^{(1)} = 0.7143, \\
x_3^{(1)} = 0.5974, & x_2^{(1)} = 0.5454, & x_1^{(1)} = 0.6062.
\end{array}$$

If we substitute these, we get, on the right side of the difference equations,

(13d)

	1	2	3	4
x_6	1	0.787	0.801	0.802
x_5	0.5	0.511	0.516	0.513
x_4	0.714	0.671	0.664	0.667
x_3	0.597	0.553	0.573	0.570
x_2	0.545	0.518	0.510	0.510
x_1	0.606	0.498	0.513	0.512 ,

and consequently new approximations. If we give the results to three places, we get

$$\text{(13e)} \qquad a_{r,r-1}y_{r-1} + a_{r,r}y_r + a_{r,r+1}y_{r+1} = a_r \,.$$

6. To treat the problem of the transverse support. Müller-Breslau has developed a *graphical method* for the solution of linear inhomogeneous difference equations of second order. Suppose the equations are

$$a_{r,r-1}y_{r-1} + a_{r,r}y_r + a_{r,r+1}y_{r+1} = a_r ;$$

and suppose that the boundary conditions $y_0 = y_{n+1} = 0$ have been given. These equations can be written in the following way:

$$(14)\qquad \frac{a_{r,r-1}\left[\dfrac{(a_{r,r-1} + a_{r,r+1})y_{r-1} + a_{r,r}y_r}{a_{r,r-1} + a_{r,r} + a_{r,r+1}}\right] + a_{r,r+1}\left[\dfrac{a_{r,r}y_r + (a_{r,r-1} + a_{r,r+1})y_{r+1}}{a_{r,r-1} + a_{r,r} + a_{r,r+1}}\right]}{a_{r,r-1} + a_{r,r+1}} = \frac{a_r}{a_{r,r-1} + a_{r,r} + a_{r,r+1}}$$

or, if we set the first bracket equal to p_r, the second to q_r, and the right side to b_r,

$$(14a)\qquad \frac{a_{r,r-1}p_r + a_{r,r+1}q_r}{a_{r,r-1} + a_{r,r+1}} = b_r .$$

Let us first assume that the y_r are known. We plot these values along the x axis parallel to one another at equal distances. In Fig. 100, let $y_r = Y_{0,r}Y_r$. If we now join Y_{r-1} with Y_r, and divide this connecting line in the ratio $a_{r,r}/(a_{r,r-1} + a_{r,r+1})$, we obtain the point P_r with the ordinate p_r. A similar construction gives the point Q_r with the ordinate q_r. Finally, if we connect P_r with Q_r and divide this line in the ratio $a_{r,r+1}/a_{r,r-1}$, we obtain a point B_r with the ordinate b_r. It is useful to draw parallels to the Y axis through the points P_r, Q_r and B_r. The first

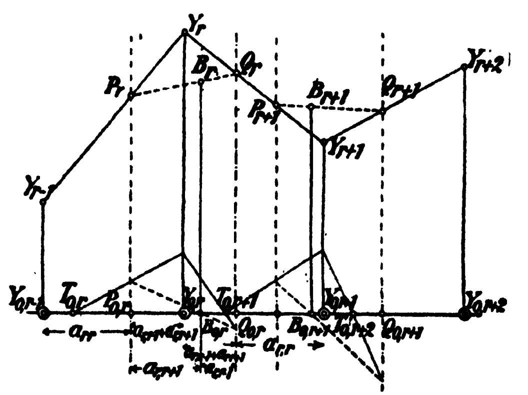


FIG. 100

two lines are dotted in Fig. 100, while the last, which is the line of division, is designated by a solid line. Since the points B_r are determined by the constants of the equation, then if we have Y_{r-1} and Y_r, we can construct Y_{r+1} in the following way. We form the intersection of $Y_{r-1}Y_r$, a line

which may be designated by g_r, with the first dotted division line at P_r. We then determine the intersection point Q_r of the connection line P_rB_r with the second dotted division line, and draw through Y_rQ_r the straight line g_{r+1} which cuts the Y parallel through $Y_{0,r+1}$ in Y_{r+1}. By a construction in the opposite direction, we can find Y_{r-1} corresponding to Y_r and Y_{r+1}. Therefore, if we have an initial side Y_0Y_1 or an end side $Y_nY_{n,+1}$ we can draw the entire Y polygon.

The following theorem is of importance in carrying out the construction. If the rth side of g_r different Y polygons passes through the fixed point T, then the neighboring sides g_{r-1} and g_{r+1} also intersect in fixed points, T_{r-1} and T_{r+1}; because the bundle of rays of the side g_r going through T cuts similar series of points on the first division line I (Fig. 101), and on the interval boundary line Y_r. The series of points of the first division line I are projected as a corresponding series of points on the second division line II by means of the pencil of rays with the center B_r. This series of points therefore corresponds to the points of the line Y_r also. The lines connecting them must therefore all pass through a point T_{r+1}, and T_r, B_r and T_{r+1} lie on a straight line. If g_r coincides with T_rB_r, then the corresponding ray of the pencil with center B_r must coincide with this line, and therefore with the line g_{r+1} also. Since all Y polygons here pass through one point because of the boundary condition $y_0 = 0$ in the first interval, they will also go through one point in the second interval, etc. Therefore it is sufficient to construct two arbitrary polygons which fulfill one boundary condition, in order to find the fixed point of the particular interval. From the ray theorems it follows that the distance of the point T_{r+1} from the limit of the interval is independent of the ordinates of the points T_r and B_r. We can therefore make these ordinates zero if only the distances from the edge of the interval are concerned. Then the point T_{r+1} will also lie on the axis. We add the index zero to these capital letters (Fig. 100) and call the polygons so determined null polygons. These are the solution polygons of the homogeneous difference equation, since $b_r = 0$. One of the null polygons is the x axis itself. A second can be constructed in the first region with arbitrarily chosen sides. This polygon $Y_{0r}|_{0\alpha}$ cuts the x axis in the fixed point $T_{0,r}$, because the x axis and all other null polygons in the first region go through the first boundary point. Also, $Y_0 = 0$ for all null polygons, because of the boundary condition. They must also go through a fixed point in the second, and consequently in the third region also, etc. The points corresponding to the right hand boundary condition can be found in the same fashion by an opposite construction. In Fig. 100, the next fixed point

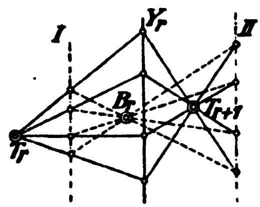


FIG. 101

$T_{0,r+1}$ is drawn from the fixed point $T_{0,r}$ by use of the division lines and the point $\bar{B}_{0r}$. We do not continue the same polygon, but, in order to keep the magnitude of the drawing within measurable limits, and to obtain good intercepts, we use a new, suitably chosen polygon through $T_{0,r+1}$ for the construction of $T_{0,r+2}$.

By use of this fixed point, the Y polygon whose corner ordinates are the solutions of the given difference equation, can easily be constructed for given B_r. For this we first draw the regions on the X axis, and construct the division lines and the reduced division line about each boundary of the region. We lay out the lengths b_r from the X axis (in a suitably chosen scale) on these lines. This gives us the points B_r. By the given construction, we find the fixed points $T_{0,r}$ and draw through them lines parallel to the Y axis. The fixed points T_r, which are the same for all Y polygons having the same left hand boundary condition, lie on these parallels. In case the first fixed point T_{01} coincides with T_1 (in Fig. 102,

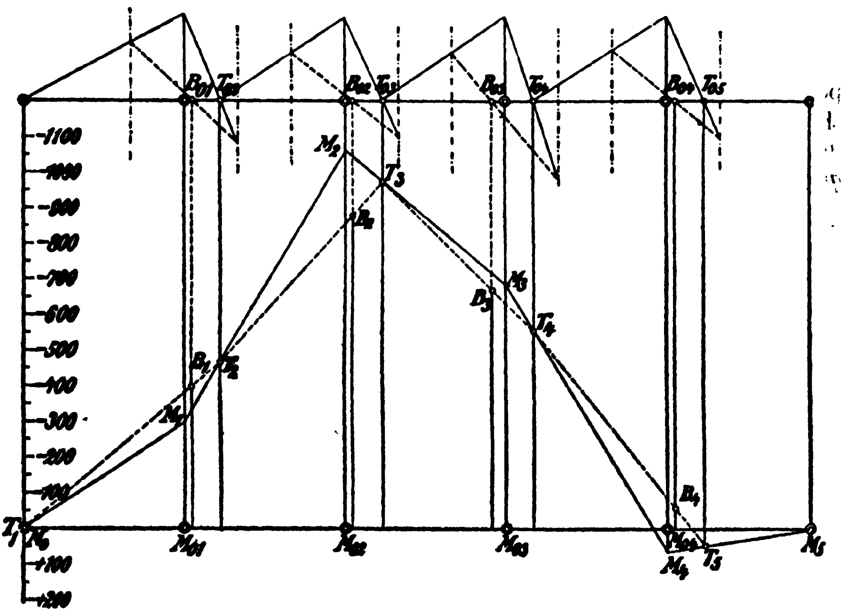


FIG. 102

the X axis is shifted downwards for the sake of greater clarity), as must be the case with the above boundary condition, then we get T_{02} as the intersection of T_1B_1 with the Y parallel through T_{02}, T_3 as the intersection of T_2B_2 with the Y parallel through T_{03}, etc. All Y polygons which satisfy the left boundary conditions must pass through these points. To satisfy the right boundary condition also, we draw the line $Y_{n+1}T_{n+1}$ out from

Y_{n+1} (denoted by M_5 in Fig. 102). This line cuts the nth boundary line at Y_n. The line Y_nT_n cuts the previous boundary line at Y_{n-1}, etc. The values Y_r thus found satisfy the difference equation with the prescribed boundary conditions. As a check the entire construction can be repeated from another boundary condition.

Fig. 102 represents the equations discussed in 23.12 in connection with the method of Massau. The equations are divided by the sum of the coefficients to put them in the proper form. This gives

$$\frac{2}{3}M_1 + \frac{4}{21}M_2 = -397\ 1/42$$

$$\frac{4}{27}M_1 + \frac{2}{3}M_2 + \frac{5}{27}M_3 = -875$$

$$\frac{5}{24}M_2 + \frac{2}{3}M_3 + \frac{1}{8}M_4 = -665\ 5/48$$

$$\frac{1}{7}M_3 + \frac{2}{3}M_4 = -54\ 1/6.$$

As solutions we get

$$M_1 = -297, \qquad M_2 = -1056, \qquad M_3 = -682, \qquad M_4 = 65$$

in good agreement with 23.12, and with the values calculated in Sec. 4 of this article.

NOTES

1. P. Funk, *Die linearen Differenzengleichungen und ihre Anwendung in der Theorie der Baukonstruktionen* (Berlin, 1920); Bleich and Melan, *Die gewöhnlichen und partiellen Differenzengleichungen der Baustatik* (Berlin, 1927).

2. Lagrange, *Nouv. Mém. Acad. sc. Berlin* **6** (1775), pp. 183-272.

3. E.g., Markoff, *Differenzenrechnung* (Leipzig, 1896); Seliwanoff, *Lehrbuch der Differenzenrechnung* (Leipzig, 1904); Nörlund, *Differenzenrechnung* (Berlin, 1924).

4. Reissner, *Archiv. für Math. u. Phys.* III, **13** (1908), pp. 317-325.

5. Hertwig has developed another process. Detailed convergence considerations, which are omitted here, are found in his work: *Festschrift für Müller-Breslau* (Leipzig, 1912), pp. 37-59.

CHAPTER FIVE

ANALYSIS OF EMPIRICAL FUNCTIONS

25. General Discussion.

1. To determine the path of a function given by observation points, we had in Chapter 2 chosen several of these data points arbitrarily, and had drawn the approximation curve or surface with these. We then obtained values for the remaining data points which agreed, more or less, with those actually given by the observations. Of course, those points used for determining the formula are reproduced exactly. These are then drawn before the others. To a certain extent they will be considered to be errorless. There is a certain arbitrariness in this procedure because the probability that data obtained empirically possess errors is in general equally large for all data.

The problem then is *to find a function which approximates all the observations as closely as possible* without favoring specific data points in the construction. This is especially desirable if we are dealing with not very precise observations of processes which, by theoretical considerations, follow simple laws. This law may then be represented by a smooth curve, while the measured data fluctuate back and forth because of errors of observation. We then seek to represent the process by a formula which corresponds to the one obtained theoretically. We assume a number of parameters in the formula, which are so determined that the curve representing the function lies as close as possible to the function under consideration (throughout the entire interval) or to the individual data points plotted.

The function obtained by this representation can then be used for finding the derivative. Otherwise this derivative could be found very approximately for such relations obtained from empirical data with rather large errors of observation.

2. There now arises the question of a *criterion for the worth of the approximation.* If only *discrete values are given,* we can observe only the behavior of the values of the approximation function under consideration corresponding to these discrete values. Nothing can be said about the behavior between these values. On the other hand, if a curve is to be approximated, then the entire course of the curve influences the approximation function. Let $y(x)$ be the curve to be approximated, and $\overline{y}(x)$ the approximation function. Then

$$\epsilon_r = \bar{y}(x_r) - y(x_r) \tag{1}$$

is the deviation for the argument value x_r. If we were to set up the condition that for discrete values the sum of all deviations, $\sum \epsilon_r = 0$, should be zero, this would not be very far reaching, because the positive and negative deviations would compensate each other. To avoid compensation of positive and negative deviations or errors, we could make the sum of the absolute values of the errors equal to zero. Mathematically, it is simpler to take the sum of an even power of the errors—actually the sum of the second powers (as was done by Gauss[1]). Naturally we do not set this sum equal to zero, but make it a minimum. In the case of discrete values then, the approximating function $\bar{y}(x_r)$ is said to be the best for which

$$M_1 = \sum (\bar{y}(x_r) - y_r)^2 = \sum \epsilon_r^2 \tag{2}$$

is a minimum. If n is the number of measurements, the mean square deviation is $\sum (\bar{y}(x_r) - y_r)^2/n$, and we can use

$$m_0 = \left(\frac{\sum (\bar{y}(x_r) - y_r)^2}{n}\right)^{1/2} \tag{3}$$

as a measure of the value of the approximation. In the method of least squares we do not use this value. Instead we employ

$$m_\kappa = \left(\frac{\sum (\bar{y}(x_r) - y_r)^2}{n - \kappa}\right)^{1/2} \tag{4}$$

where κ is the value of the parameter to be determined suitably. This value m is known as the *mean error* of the individual measurement.[2]

In the same way, when a *continuous curve* $y(x)$ *is being considered,* we shall not use the integral of the difference between the ordinates of the given and the approximating curves, but an integral whose integrand is the square of this difference of ordinates. Then the best approximation is the one for which

$$M_2 = \int_a^b (\bar{y}(x) - y(x))^2 \, dx \tag{5}$$

is a minimum. We use the square root of the mean square of the deviations

$$m = \left(\frac{\int_a^b (\bar{y}(x) - y(x))^2 \, dx}{b - a}\right)^{1/2} \tag{6}$$

as a measure of the worth of the approximation.

3. We shall limit ourselves here to the representation of *the relation by a linear combination of suitably chosen functions.* Therefore we write

$$\bar{y}(x) = c_0g_0(x) + c_1g_1(x) + c_2g_2(x) + \cdots + c_ng_n(x). \tag{7}$$

The constants c_m are then so determined that the value of the expression for M_1 or M_2, given above, is a minimum; i.e., for discrete values,

$$M_1 = \sum_1^s (y_r - c_0g_0(x_r) - c_1g_1(x_r) - \cdots - c_ng_n(x_r))^2 = \sum_1^s \epsilon^2 \tag{8}$$

is a minimum, while for continuous functions,

$$M_2 = \int_a^b (y(x) - c_0g_0(x) - c_1g_1(x) - \cdots - c_ng_n(x))^2\, dx \tag{9}$$

is a minimum.

In this way the difference between the two functions $y(x)$ and $\bar{y}(x)$ will be small throughout the entire interval; i.e., an approximation is obtained whose value is estimated according to its behavior throughout the whole interval.

4. To make M a minimum, we have at our disposal only the coefficients c_m because no parameters are contained in the functions $g_m(x)$ or, if there were such parameters, they would have been suitably chosen beforehand. The expression M may then be considered as a function of the coefficients c_m. According to well-known theorems, a necessary *condition that M be a minimum* is that

$$\frac{\partial M}{\partial c_m} = 0 \tag{10}$$

for all values of m. For discrete observation values, this gives the $n + 1$ linear "*normal equations*"

$$\begin{aligned}
\frac{1}{2}\frac{\partial M}{\partial c_0} &= c_0[g_0^2] + c_1[g_0g_1] + \cdots + c_n[g_0g_n] - [yg_0] = 0,\\
\frac{1}{2}\frac{\partial M}{\partial c_1} &= c_0[g_0g_1] + c_1[g_1^2] + \cdots + c_n[g_1g_n] - [yg_1] = 0,\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
\frac{1}{2}\frac{\partial M}{\partial c_n} &= c_0[g_0g_n] + c_1[g_1g_n] + \cdots + c_n[g_n^2] - [yg_n] = 0,
\end{aligned} \tag{11}$$

where the more convenient Gaussian bracket notation is used in place of the summation sign $\sum$. For the continuous case, integrals replace the sums, so that the normal equations become:

$$(12)\quad \begin{aligned}
\frac{1}{2}\frac{\partial M}{\partial c_0} &= c_0 \int_a^b (g_0(x))^2\,dx + c_1 \int_a^b g_0(x)g_1(x)\,dx + \cdots \\
&\qquad + c_n \int_a^b g_0(x)g_n(x)\,dx - \int_a^b yg_0(x)\,dx = 0, \\
\frac{1}{2}\frac{\partial M}{\partial c_1} &= c_0 \int_a^b g_1(x)g_0(x)\,dx + c_1 \int_a^b (g_1(x))^2\,dx + \cdots \\
&\qquad + c_n \int_a^b g_1(x)g_n(x)\,dx - \int_a^b yg_1(x)\,dx = 0, \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
\frac{1}{2}\frac{\partial M}{\partial c_n} &= c_0 \int_a^b g_n(x)g_0(x)\,dx + c_1 \int_a^b g_n(x)g_1(x)\,dx + \cdots \\
&\qquad + c_n \int_a^b (g_n(x))^2\,dx - \int_a^b yg_n(x)\,dx = 0.
\end{aligned}$$

The coefficients can be calculated from the normal equations provided that the determinant is not zero.

In approximating the path of a function by such a linear combination of functions, it is not necessary that the sum of the errors (or the integral over the error surface)

$$(13)\qquad [\bar{y}_r - y_r] \qquad \text{or} \qquad \int_a^b (\bar{y} - y)\,dx$$

be zero. In general this is the case only if an additive constant appears in the series of the approximation functions, if, therefore, we set $g_0(x) = 1$. The first normal equation then takes the form (13).

5. The expression for M can be transformed still further. If we take the case of discrete observations, then

$$(14)\qquad M_1 = [\epsilon\epsilon] = [(\bar{y}(x_r) - y_r)^2] = [\bar{y}^2] - 2[y\bar{y}] + [y^2].$$

If we now multiply the corresponding normal equations (11) by the coefficients c_0, c_1, $\cdots$, c_n and add, we get

$$(15)\quad \begin{aligned}
\frac{1}{2}\left(c_0\frac{\partial M}{\partial c_0} + c_1\frac{\partial M}{\partial c_1} + \cdots + c_n\frac{\partial M}{\partial c_n}\right) &= c_0[g_0\bar{y}] + c_1[g_1\bar{y}] + \cdots \\
&\qquad + c_n[g_n\bar{y}] - [y\bar{y}] = [\bar{y}^2] - [y\bar{y}] = 0.
\end{aligned}$$

Therefore, by means of the relation derived above, we can put the minimum value of the sum of the squares of the errors in the two forms

$$M_1^0 = [\epsilon\epsilon] = [y^2] - [y\bar{y}] = [y^2] - [\bar{y}^2]. \tag{16}$$

For continuous functions we find

$$M_2^0 = \int_a^b y^2(x)\,dx - \int_a^b y(x)\bar{y}(x)\,dx = \int_a^b y^2(x)\,dx - \int_a^b \bar{y}^2(x)\,dx \tag{17}$$

in the same way. If we write out the first form of the equations,

$$-c_0[yg_0] - c_1[yg_1] - \cdots - c_n[yy_n] + [yy] = [\epsilon\epsilon] = M_1^0\,, \tag{18}$$

$$\begin{aligned} &- c_0 \int_a^b yg_0(x)\,dx - c_1 \int_a^b yg_1(x)\,dx - \cdots \\ &\qquad - c_n \int_a^b yg_n(x)\,dx + \int_a^b y^2(x)\,dx = M_2^0\,, \end{aligned} \tag{19}$$

and add this equation as the $(n + 2)$nd to the corresponding set of normal equations (11) or (12), we have a system of $n + 2$ equations, from which we can calculate the $n + 1$ unknowns and the value of M, according to the scheme given in 23.3. Since we are dealing with symmetric equations, we obtain the part of the scheme lying above the dotted line.

NOTES

1. Gauss, *Abhandlungen zur Methode der kleinsten Quadrate*. Edited by Börsch and Simon (Berlin, 1887).

2. Cf. *Enzyklopädie d. math. Wissenschaft* I, D, 2; Bauschinger, *Ausgleichungsrechnung*, or any other book on approximation of functions.

26. Approximation by Linear Functions.

1. The finding of the approximation function becomes especially simple if we are dealing with functions in which the *one variable is a linear* (*or approximately linear*) *function of the other*, so that approximation by a linear function is sufficient. When such observations are plotted on ordinary coordinate paper, deviations from the straight line occur, either because of the errors of observation or for other reasons. The problem is to find the straight line best representing the relation.

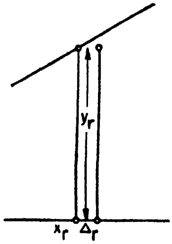


FIG. 103

If we are concerned with discrete observations, then the argument values x_r, as well as the corresponding function values, can be measured incorrectly. However, we shall consider the argument values, plotted as abscissa values, as exact, and ascribe the errors to the function values, plotted as ordinates. For example, if we have measured the correct ordinate value y_r to an inaccurately measured abscissa value x_r, for which we have measured $x_r + \Delta_r$, then we take the abscissa value $x_r + \Delta_r$ as correct and assign to it the now inaccurate ordinate value y_r.

If y_r are the measured values, $\bar{y}_r$ the desired approximation value, then, by 25(8), the sum of the squares of the errors,

$$M_1 = \sum_1^n \Delta_r^2 = \sum_1^n (\bar{y}_r - y_r)^2, \tag{1}$$

must be a minimum. If $\bar{y}$ is a linear function

$$\bar{y} = ax + b, \tag{1a}$$

then the error of the individual measurement is $\Delta_r = y_r - ax_r - b$. Therefore in

$$M_1 = \sum_1^n (y_r - ax_r - b)^2, \tag{2}$$

the parameters a and b are so chosen that this expression becomes a minimum; i.e., by 25(11),

$$\begin{aligned} \frac{1}{2}\frac{\partial}{\partial a}\sum_1^n (y_r - ax_r - b)^2 &= -\sum_1^n (y_r - ax_r - b)x_r = 0, \\ \frac{1}{2}\frac{\partial}{\partial b}\sum_1^n (y_r - ax_r - b)^2 &= -\sum_1^n (y_r - ax_r - b) = 0 \end{aligned} \tag{3}$$

or, if we again use the Gaussian notation, the normal equations become

$$[y_r x_r] - a[x_r^2] - b[x_r] = 0, \qquad [y_r] - a[x_r] - bn = 0. \tag{4}$$

From this it follows that

$$b = \frac{[y_r][x_r^2] - [x_r y_r][x_r]}{n[x_r^2] - [x_r]^2}, \qquad a = \frac{n[x_r y_r] - [x_r][y_r]}{n[x_r^2] - [x_r]^2}. \tag{5}$$

2. According to Mehmke,[3] the *approximating straight line can be obtained graphically* in the following manner. The second of both normal equations can be written in the form

$$\frac{[y_r]}{n} = a\frac{[x_r]}{n} + b; \qquad y_s = ax_s + b, \tag{6}$$

where $y_s = [y_r]/n$, $x_s = [x_r]/n$ are the coordinates of the center of mass of the points plotted. Therefore the approximating straight line, according to the second equation, must be a line through this center of mass.

After the center of mass S has been ascertained, we draw a number of lines through it which could satisfy the problem. We then measure for these the deviations of the measured points, in the direction of the ordinate axis. We calculate $\sum \Delta_r^2$ and plot this value, in arbitrary scale units, perpendicular to some auxiliary line at its intersection with the straight lines of the pencil through the center of mass. This is done for all the chosen approximating lines, and the points which are obtained are then con-

nected by a smooth curve. This curve will have a minimum. If we drop a perpendicular from this point to the auxiliary line, then we have a second point of the desired approximation line.

According to Werkmeister,[4] it is simpler to find a second point of the line if we assign to each point the weight x_r. Then the coordinates of this center of mass $x_\sigma = ([x_r^2])/[x_r])$, $y_\sigma = ([y_r x_r]/[x_r])$ satisfy the first equation. Therefore the approximation line must also pass through this point $\sum$. Consequently we have two points of the line. The summations can be carried out numerically or graphically.

3. *Example*: Perry[5] gives in his work on applied mechanics the energy loss of a pump in mkg as a function of the height h:

h in m	6	15	30	60	90	120
Q in mkg	10.25	10.42	11.26	12.52	13.78	15.04

The dependence is to be represented by a linear function.

If we want to make the approximation graphically, we must first

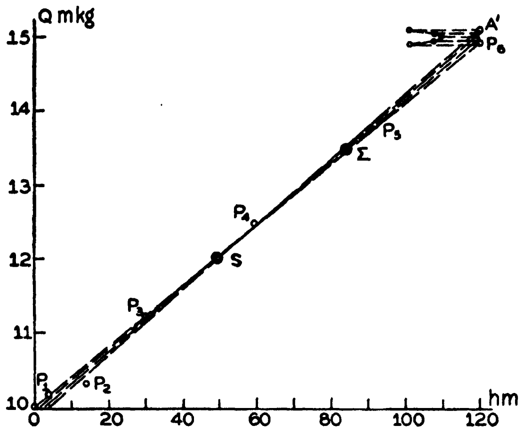


FIG. 104

determine the coordinate s of the center of mass from these numerical values, by graphical or numerical addition, and division by 6:

$$h_s = 53.5\text{m.},\qquad Q_s = 12.212 \text{ mkg}.$$

The desired line is then sA', by the method of Mehmke. This has the equation

$$Q = 9.92 + 0.0428h.$$

For the coordinates of the second center of mass, we find

$$h_\sigma = 84.9,\qquad Q_\sigma = 13.557.$$

This gives the point $\sum$. The straight line $s\sum$ coincides (in Fig. 104) with sA', as it indeed must.

If we want to carry out the operation numerically, then it is advisable first to assume an approximation line, and then to set up the normal equations for the improvement of the approximation values thus obtained. Let us take as the first approximation the line through the third and sixth points,

$$\frac{\bar{Q} - 11.26}{h - 30} = \frac{3.78}{90}; \qquad \bar{Q} = 0.042h + 10 = \bar{a}h + \bar{b}. \tag{7}$$

We get approximation values for $\bar{Q}$ from these approximation values $\bar{a}$ and $\bar{b}$. To obtain the correct values, we add the correction ΔQ, Δa and Δb to these. The deviations of the observed points of the approximation curve are

$$\Delta_r = \bar{Q}_r + \Delta Q_r - (\bar{a} + \Delta a)h_r - (\bar{b} + \Delta b) = \Delta Q_r - \Delta a h_r - \Delta b \tag{7a}$$

if we apply equation (7). Then

$$\sum(\Delta Q_r - \Delta a h_r - \Delta b)^2 \tag{8}$$

is to be made a minimum. The numbers appearing in the formation of the normal equations are now all so small that we can carry out the multiplications mentally, or with a slide rule. We get

(8a)

$[h]$	Q	$\bar{Q}$	$[\Delta Q]$	$[h\Delta Q]$	$[h^2]$
6	10.25	10.252	−0.002	−0.012	36
15	10.42	10.62	−0.210	−3.150	225
30	11.26	11.26	0.00	0.00	900
60	12.52	12.52	0.00	0.00	3600
90	13.78	13.78	0.00	0.00	8100
120	15.04	15.04	0.00	0.00	14400
321			−0.212	−3.162	27261.

From this we obtain

(8b)

$$\Delta a = \frac{6 \cdot 3.162 - 0.212 \cdot 321}{321^2 - 6 \cdot 27261} = \frac{18.97 - 68.05}{103041 - 163566} = \frac{49.08}{60525} = 0.00081,$$

$$\Delta b = \frac{0.212 \cdot 27261 - 3.162 \cdot 321}{321^2 - 6 \cdot 27261} = \frac{5779 - 1015}{-60525} = -\frac{4764}{60525} = -0.0787,$$

by the use of equation (5); i.e., we get as the equation of the approximation line

$$Q = 0.04281h + 9.9213, \tag{8c}$$

which is in good agreement with the result found graphically.

If a continuous curve is to be approximated by a straight line, then we must replace the sums by the corresponding integrals. These integrals will perhaps be evaluated graphically. Otherwise, we proceed exactly as above.

4. If we are to plot the measurements on *function scales,* rather than on coordinates with uniform scales, and if these are so chosen that the path of the function can be represented by a straight line, then the above construction needs an alteration, given by *Schwerdt.*[6] Suppose the function scales on the respective axes are

$$\xi = \varphi(x), \qquad \eta = \psi(y). \tag{9}$$

Just as with uniform scales, we now assume, for the case of discrete observations, that the argument values x_r, and therefore the abscissas ξ_r, are correct. Then the observation errors are transferred to the function values y_r and hence to the ordinates η_r. Therefore to the abscissa $\xi_r = \varphi(x_r)$ is plotted not the correct ordinate $\bar{\eta}_r = \psi(\bar{y}_r)$ but the inaccurate $\eta_r = \psi(y_r)$ which has the error

$$\Delta\eta_r = \psi(y_r) - \psi(\bar{y}_r) \approx \psi'(\bar{y}_r)\Delta y_r \approx \psi'(y_r)\Delta y_r\,. \tag{10}$$

The equation of the approximation function is now

$$\bar{\eta} = a\xi + b. \tag{11}$$

i.e., the error is

$$\Delta\eta_r = \eta_r - a\xi_r - b.$$

We then get

$$\Delta y_r = \frac{\Delta\eta_r}{\psi'(y_r)} = \frac{\eta_r}{\psi'(y_r)} - a\frac{\xi_r}{\psi'(y_r)} - \frac{b}{\psi'(y_r)} \tag{13}$$

for the error of an individual measurement. From the condition that the sum of the squares of the errors $\sum (\Delta y_r)^2$ be a minimum, we get the two normal equations for the determination of a and b:

$$\left[\frac{\eta}{\psi'^2}\right] - a\left[\frac{\xi}{\psi'^2}\right] - b\left[\frac{1}{\psi'^2}\right] = 0, \left[\frac{\eta\xi}{\psi'^2}\right] - a\left[\frac{\xi^2}{\psi'^2}\right] - b\left[\frac{\xi}{\psi'^2}\right] = 0. \tag{14}$$

The approximating line can be found graphically as above. We need only assign to the rth point the weight $1/(\psi_r')^2$. Then the coordinates of the center of mass are

(15) $$\eta_s = \left[\frac{\eta}{\psi'^2}\right] : \left[\frac{1}{\psi'^2}\right]; \qquad \xi_s = \left[\frac{\xi}{\psi'^2}\right] : \left[\frac{1}{\psi'^2}\right].$$

By division with $[1/(\psi)')^2]$, the first normal equation (14) can then be written

(15a) $$\eta_s - a\xi_s - b = 0.$$

The approximation line therefore passes through the center of mass thus determined. We proceed exactly as above, except that we either plot $[(\Delta\eta_r)^2/(\psi')^2]$ perpendicularly at the various points of intersection with the auxiliary lines, and determine the second point of the approximation line by the minimum of this expression, or, more simply, we attribute the weight $\xi_r/(\psi_r')^2$ to each point here also. Then we get a second center of mass with coordinates

(15b) $$\eta_\sigma = \left[\frac{\eta\xi}{\psi'^2}\right] : \left[\frac{\xi}{\psi'^2}\right]; \qquad \xi_\sigma = \left[\frac{\xi^2}{\psi'^2}\right] : \left[\frac{\xi}{\psi'^2}\right].$$

By the second equation, these coordinates also satisfy the equations of the approximation line. Hence this line is determined if the two points are not too close together.

5. *Example*: The experiments of Fuhrmann[7] on the air resistance w of his balloon model V can be represented by an equation of the form $w = Av^n$, where v is the air velocity in meters per second, A and n are constants. From the observations, we take the following data:

v	1	2	3	4	5	6	7	8	9	10m
w	0.2	0.8	1.5	2.7	3.7	5.2	6.7	8.3	9.9	11.5kg.

If we put the equation $w = Av^n$ in logarithmic form, and write $\eta = \log w$, $\xi = \log v$, then we obtain (if we plot the data on double logarithmic paper) an approximation

$$\eta = a\xi + b.$$

The weight of the individual point is $1/(\psi')^2 = w^2(\ln 10)^2$ but since the constant factor $(\ln 10)^2$ is canceled out by division, we can omit this factor. From the scheme of the next section, in which the problem is handled numerically, we take

$$\left[\frac{\xi}{\psi'^2}\right] = 362.18, \quad \left[\frac{\eta}{(\psi')^2}\right] = 368.85, \quad \left[\frac{1}{(\psi')^2}\right] = 394.99,$$

so that we find for the coordinates of the center of mass S

$$\xi_s = 0.918, \qquad \eta_s = 0.934.$$

The second point A' of the approximation line is found either by the

minimum of the quantity $\sum (\Delta\eta_r/\psi'_r)^2$ for the lines plotted experimentally (which are not drawn in Fig. 105), or by the calculation of

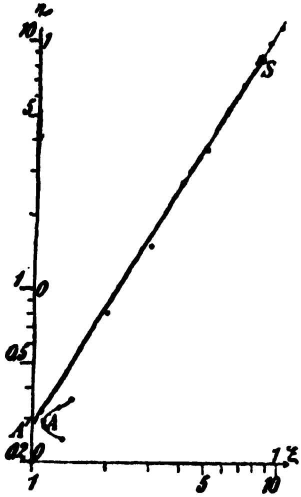


FIG. 105

$$\left[\frac{\xi\eta}{(\psi')^2}\right] = 344.48, \qquad \left[\frac{\xi^2}{(\psi')^2}\right] = 336.$$

From this it follows that $\xi_s = 0.927$, $\eta_s = 0.952$. These points lie too close together for us to draw a line through them with any certainty.

From the drawing we get

$$a = 1.610,$$

$$b = -0.542 = 0.458 - 1$$

$$= \log A = \log 0.287,$$

so that we find as an approximation function

$$w = 0.287v^{1.610}.$$

6. We shall now carry out the *numerical approximation* of the above example, without previous formulation of an approximating function. From the table given here we see that we are dealing with materially larger numbers than in the use of a preliminary approximating function.

	ψ	η	$[1/\psi'^2]$	$[\xi/\psi'^2]$	$[\eta/\psi'^2]$	$[\xi\eta/\psi'^2]$	$[\xi^2/\psi'^2]$
	0	−0.6990	0.04	0	−0.03	0	0
	0.3010	−0.0969	0.64	0.19	−0.06	−0.02	0.06
	0.4771	0.1761	2.25	1.07	+0.40	+0.19	0.51
	0.6021	0.4314	7.29	4.39	3.14	1.89	2.64
	0.6990	0.5682	13.69	9.57	7.78	5.44	6.69
	0.7782	0.7160	27.04	21.04	19.36	15.07	16.37
	0.8451	0.8261	44.89	37.94	37.08	31.34	32.06
	0.9031	0.9191	68.89	62.21	63.32	57.18	56.18
	0.9542	0.9956	98.01	93.52	97.58	93.11	89.24
	1.0000	1.0607	132.25	132.25	140.28	140.28	132.25
Σ			394.99	362.18	368.85	344.48	336.00.

By (15) this becomes

$$a = \frac{344.48 \times 394.99 - 368.85 \times 362.18}{336 \times 394.99 - (362.18)^2} = 1.606,$$

$$b = \frac{368.85 \times 336 - 344.48 \times 362.18}{336 \times 394.99 - (362.18)^2} = -0.538 = \log 0.290.$$

Therefore, this gives as the approximating function

$$w = 0.290 v^{1.606},$$

which is in close agreement with the equation found graphically.

Rothe[6] gives another example which also can be treated very well graphically: The axes of a rotation ellipsoid are to be determined which best approximate a rotation body of approximately this form. Corresponding values of x and r are determined with a micrometer. These values are plotted by placing the quadratic scales $\xi = x^2$, $\eta = r^2$ on the ordinates, and the plotted points are approximated by a straight line, according to the method above. The intercepts of the straight lines on the ordinates permit the lengths of the axes to be read off directly on the quadratic scales.

NOTES

1. Schwerdt, *Lehrbuch der Nomographie* (Berlin, 1924), p. 82.
2. Werkmeister, *Z. f. angew. Math. u. Mech.* I (1921), p. 491.
3. Perry, *Angewandte Mechanik* (Berlin, 1908), p. 513.
4. Schwerdt, *op. cit.*, p. 84.
5. Fuhrmann, *Z. f. Flugtechnik* 2 (1911), pp. 165-167.
6. Rothe, *Elektrotechnische Zeitschrift*, 41 (1920), pp. 999-1002.

27. Approximation by Rational Integral Functions.

1. In cases in which we know nothing about the path of the function, or in which we make an approximate numerical calculation of complicated functions, not already tabulated, it is most convenient to approximate the functions by a power series. We therefore set

$$g_0(x) = 1, \qquad g_1(x) = x, \qquad \cdots, \qquad g_n(x) = x^n.$$

We introduce the limits -1 and $+1$, in place of the interval limits a and b, by the substitution

$$x = \frac{a+b}{2} - \frac{a-b}{2}\,t. \tag{1}$$

If we set up the approximating function

$$\bar{y} = a_0 + a_1 t + a_2 t^2 + \cdots + a_n t^n, \tag{2}$$

then, by 25(11), the normal equations become

$$
(3)\qquad
\begin{aligned}
&na_0 + [t]a_1 \quad + [t^2]a_2 \cdots [t^n]a_n \quad - [y] \quad = 0,\\
&[t]a_0 + [t^2]a_1 \quad + [t^3]a_2 \cdots [t^{n+1}]a_n \; - [yt] \quad = 0,\\
&\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot\\
&[t^n]a_0 + [t^{n+1}]a_1 + [t^{n+2}]a_2 \cdots [t^{2n}]a_n - [yt^n] = 0.
\end{aligned}
$$

As the $(n + 2)$nd equation, we take the equation 25(18) for the calculation of the *sum of the squares of the errors*:

$$
(4)\qquad -[y]a_0 - [yt]a_1 - [yt^2]a_2 - [yt^n]a_n + [yy] = M^0.
$$

Consequently we have a system of symmetric equations of the form considered in 23.3, from which we can calculate the unknowns $a_0 \cdots a_n$ by the scheme on page 270. In addition we can determine the value M_0 , and therefore the sum of the squares of the errors. If the measurements are made for equidistant values of the argument, then the sum of the odd powers of t becomes zero. Therefore the *normal equations divide into two groups*. From the one group we can calculate the coefficients with even index, and the corresponding terms of the sum of the squares of the errors, and from the other group those coefficients with odd index.

Nothing can be said as a general rule about the degree of the approximating function to be selected. The best we can do is to plot the measured points on rectangular coordinate paper, draw a smooth curve through them, and consider what must be the degree of the curve with which we can best represent this curve. Naturally we choose as low a degree as possible for the approximating curve.

2. *Example*: In the *Lehrbuch der Physik* of Muller-Pouillet, the dependence of the specific heat of water c on the temperature θ is given by the following data:

t	θ	c	y	t	θ	c	y
−1	0	1.0075	7.5	0.2	60	0.9995	−0.5
−0.8	10	1.0008	0.8	0.4	70	1.0012	+1.2
−0.6	20	0.9974	−2.6	0.6	80	1.0032	+3.2
−0.4	30	0.9971	−2.9	0.8	90	1.0057	+5.7
−0.2	40	0.9974	−2.6	1	100	1.0086	+8.6.
0	50	0.9983	−1.7				

From the curve drawn in Fig. 106, we see that, because of the asymmetric path, a function of at least third degree must be used for the approximation.

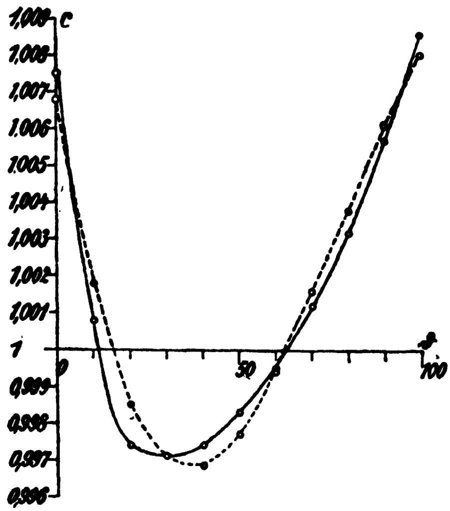


FIG. 106

First we introduce the limits ± 1 by the substitution

$$\theta = 50(t + 1)$$

and to get smaller numbers, we set $y = (c - 1)10.^3$ The normal equations can be calculated from the data:

$$\begin{aligned}
11a_0 \qquad + 4.4a_2 \qquad\qquad - 16.7 &= 0 \\
4.4a_1 \qquad + 3.1328a_3 - 10.56 &= 0 \\
4.4a_0 \qquad + 3.1328a_2 \qquad\qquad - 20.08 &= 0 \\
3.1328a_1 \qquad + 2.6259a_3 - 5.1408 &= 0.
\end{aligned}$$

Therefore there are two groups of equations which are, with the addition of the corresponding parts of the expression for $M_0 - [yy]$,

$$\begin{aligned}
11a_0 + 4.4a_2 - 16.7 &= 0 \\
4.4a_0 + 3.1328a_2 - 20.08 &= 0 \\
-16.7a_0 - 20.08a_2 + 0 &= A_1 \,,
\end{aligned}$$

$$\begin{aligned}
4.4a_1 + 3.1328a_3 - 10.56 &= 0 \\
3.1328a_1 + 2.6259a_3 - 5.1408 &= 0 \\
-10.56a_1 - 5.1408a_3 + 0 &= A_2 \,.
\end{aligned}$$

If we use the scheme given in 23.3, we get for the coefficients,

$$a_0 = -2.386 \qquad a_2 = 9.761$$

$$a_1 = 6.682. \qquad a_3 = -6.014.$$

If we again divide by 1000, we get

$$c = 1 - 0.002386 + 0.006682t + 0.009761t^2 - 0.006014t^3,$$

and since $[yy] = 200.09$, this becomes

$$M_0 = 200.09 \times 10^{-6} - 156.15 \times 10^{-6} - 39.65 \times 10^{-6} = 4.29 \times 10^{-6}.$$

Therefore, the mean error of an observation is, by 25(4),

$$m = \left(\frac{4.29 \times 10^{-6}}{11 - 4}\right)^{1/2} = 0.78 \times 10^{-3}.$$

For coefficients therefore, we can carry along at most four decimal places, i.e.,

$$c = 0.9977 + 0.0067t + 0.0098t^2 - 0.0060t^3.$$

The curve which is obtained is dotted in Fig. 106. If we form the sum of the squares of the errors from the given values and the eleven values plotted in the figure, we get

$$[\epsilon\epsilon] = 4.29 \times 10^{-6},$$

in complete agreement with the value calculated above for M_0. For the original variables we get

$$c = 1.0068 - 0.00026\theta + 0.0000112\theta^2 - 0.000000048\theta^3.$$

If the approximation obtained is not sufficient, we can approximate by a rational integral function of fourth degree. Then the system of normal equations for the odd coefficients remains the same, while the calculation for the even coefficients must be repeated.

3. For *the approximation curve of widely scattered observations,* we are usually satisfied to make a graphical construction. We plot the observations on rectangular coordinate paper and draw a *smooth curve* amongst these points, i.e., a curve for which at least the first derivative is continuous.

For a check on the path of this curve, we can make use of the approximation by rational integral functions is still another way, if the observations are equidistant.

First we assume that the curvature changes so slowly that, with sufficient accuracy, we can approximate the portion connecting any three neighboring points with a straight line. By the coordinate transformation $t = (x - x_0)/h$ the argument values $x_{-1} = x_0 - h$, x_0, $x_{+1} = x_0 + h$ of these three points are assigned the values -1, 0, $+1$ of t. If the approximation line has the equation $\bar{y} = a_0 + a_1 t$, then

$$\sum_{-1}^{+1} (y - a_0 - a_1 t)^2 \tag{5}$$

must be a minimum; i.e.,

$$\begin{aligned} &\sum_{-1}^{+1} (y - a_0 - a_1 t) = y_{-1} + y_0 + y_{+1} - 3a_0 = 0, \\ &\sum_{-1}^{+1} (y - a_0 - a_1 t)t = -y_{-1} + y_{+1} - 2a_1 = 0. \end{aligned} \tag{6}$$

We are only interested in improving the middle ordinate. This becomes

$$\bar{y}_0 = a_0 = \frac{1}{3}(y_{-1} + y_0 + y_{+1}). \tag{7}$$

This is the ordinate of the center of mass of the three points. Neglecting one of the points, we connect the other two P_{m-1}, P_{m+1} by a straight line, which cuts the ordinate of the omitted point P_m at Q_m (Fig. 107). The length $Q_m P_m$ is then divided into three equal parts, and the division point $\bar{P}_m$ adjacent to the point Q_m is chosen as the new approximation point. Then all the points t, except the two end points, can be corrected. If the points so obtained are still widely scattered, the process can be repeated.

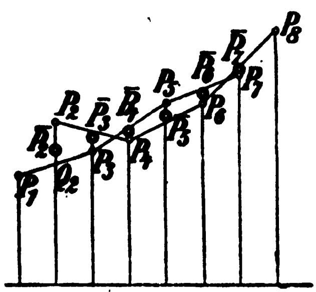


FIG. 107

4. If the *curvature changes too rapidly*, then we no longer make approximations with segments of straight lines, but use *parabolas* instead. To construct these, we use the point itself and the four neighboring points. We therefore use the points which have the abscissas -2, -1, 0, $+1$, $+2$ and the ordinates y_{-2}, y_{-1}, y_0, y_{+1}, y_{+2} after the above transformation. Let the equation of the approximation parabola be

$$\bar{y} = a_0 + a_1 t + a_2 t^2. \tag{8}$$

Then

$$\sum_{-2}^{+2} (y - a_0 - a_1 t - a_2 t^2)^2 \tag{9}$$

must be a minimum. This gives the three normal equations

$$\begin{aligned}
&\sum_{-2}^{+2} (y - a_0 - a_1 t - a_2 t^2) = \sum_{-2}^{+2} y - 5a_0 - 10a_2 = 0, \\
(10) \quad &\sum_{-2}^{+2} (yt - a_0 t - a_1 t^2 - a_2 t^3) = \sum_{-2}^{+2} ty - 10a_1 = 0, \\
&\sum_{-2}^{+2} (yt^2 - a_0 t^2 - a_1 t^3 - a_2 t^4) = \sum_{-2}^{+2} t^2 y - 10a_0 - 34a_2 = 0,
\end{aligned}$$

if we observe that $\sum_{-2}^{+2} t = 0$, $\sum_{-2}^{+2} t^2 = 10$, $\sum_{-2}^{+2} t^3 = 0$ and $\sum_{-2}^{+2} t^4 = 34$. Since we are only concerned with the corrected value $\bar{y}_0 = a_0$ which we obtain for $t = 0$, then we calculate this value by elimination of a_2 from the first and third equations:

$$\begin{aligned}
35a_0 &= 17 \sum_{-2}^{+2} y - 5 \sum t^2 y \\
&= -3y_{-2} + 12y_{-1} + 17y_0 + 12y_{+1} - 3y_2 \, .
\end{aligned} \tag{11}$$

Now by 10(6),

$$\Delta_0^4 = y_{-2} - 4y_{-1} + 6y_0 - 4y_1 + y_2 \, . \tag{12}$$

If we introduce this, we get

$$35a_0 = -3\Delta_0^4 + 35y_0 \, , \tag{13}$$

or

$$\bar{y}_0 = a_0 = y_0 - \frac{3}{35} \Delta_0^4 \, . \tag{14}$$

In general we shall be able to draw a smooth curve through the resultant points after a single correction. If this is not the case, then the same process can be repeated, until the corrections which appear become vanishingly small in comparison to the ordinates. Because of the vanishing of Δ^4, we have then made a stepwise approximation by a function of third degree. The disadvantage of this method, in comparison to those given in Sec. 1 of this article, is that we can get no improvements on the first and second, and the last and next to last ordinates, because we cannot form the corresponding differences of fourth order.[1]

5. *Example*: If we investigate the elastic hysteresis loop of steel bars under various large loadings and unloadings, for example, for the load σ and extension ϵ, we find the values given in the first two columns:[2]

σ kg	ϵ	Δ^1	Δ^2	Δ^3	Δ^4	$-3/35\ \Delta^4$	$\bar{\epsilon}$	$-3/35\ \bar{\Delta}^4$	$\bar{\bar{\epsilon}}$
900	5491						5491		5491
		103							
1000	594		1				594		594
		104		+ 5					
1100	698		6		− 6	+0.5	698.5	+0.2	698.7
		110		− 1					
1200	808		5		+ 2	−0.2	807.8	−0.2	807.6
		115		+ 1					
1300	923		6		+ 6	−0.5	922.5	+0.3	922.8
		121		+ 7					
1400	6044		13		−16	+1.4	6045.4	−0.4	6045
		134		− 9					
1500	178		4		+30	−2.6	175.4	−0.7	174.7
		138		+21					
1600	316		25				316		316
		163							
1700	479						479		479

The difference scheme is formed from the values of ϵ, $-3/(35)\Delta^4$ is calculated, and this correction is added to the value ϵ giving $\bar{\epsilon}$. The same process can again be applied to these values. The second corrections are generally much smaller.

6. If we are not dealing with discrete observations, but with continuous curves, then integrals replace the sums. The integrals, after the transformation given in Sec. 1, take the form

$$(15) \qquad \int_{-1}^{+1} g_m(t)\cdot g_n(t)\,dt = \int_{-1}^{+1} t^{m+n}\,dt = \begin{cases} \dfrac{2}{m+n+1} & \text{if } m+n \text{ is odd} \\ 0 & \text{if } m+n \text{ is even} \end{cases}$$

while we write for the constant terms,

$$(16)\ 2J_0 = \int_{-1}^{+1} dt;\quad 2J_1 = \int_{-1}^{+1} y\cdot t\,dt;\quad \cdots;\quad 2J_m = \int_{-1}^{+1} y\cdot t^m\,dt.$$

If we represent the coefficients of the approximating function with Q_m as above, then the normal equations 25(12) become, upon division by 2,

$$
\begin{array}{llllll}
 & a_0 & + \frac{1}{3} a_2 & + \frac{1}{5} a_4 & \cdots - J_0 = 0 \\
 & \frac{1}{3} a_1 & + \frac{1}{5} a_3 & + \frac{1}{7} a_5 & \cdots - J_1 = 0 \\
(17) & \frac{1}{3} a_0 & + \frac{1}{5} a_2 & + \frac{1}{7} a_4 & \cdots - J_2 = 0 \\
 & \frac{1}{5} a_1 & + \frac{1}{7} a_3 & + \frac{1}{9} a_5 & \cdots - J_3 = 0 \\
 & \frac{1}{5} a_0 & + \frac{1}{7} a_2 & + \frac{1}{9} a_4 & \cdots - J_4 = 0 \\
 & \cdots & \cdots & \cdots & \cdots
\end{array}
$$

The equations are again divided into two groups; we can calculate the coefficients with even index from the first and those with odd index from the second. If we add the corresponding terms from the equation for the square of the errors (25(19)),

$$
(18) \quad C = \frac{1}{2} M_2^0 - \frac{1}{2} \int_{-1}^{+1} y^2 \, dt = -a_0 J_0 - a_1 J_1 - a_2 J_2 - \cdots - a_n J_n \, ,
$$

then we get, for example for $n = 4$, the two groups

$$
\begin{array}{lll}
 & a_0 + \frac{1}{3} a_2 + \frac{1}{5} a_4 - J_0 = 0 & \frac{1}{3} a_1 + \frac{1}{5} a_3 - J_1 = 0 \\
 & \frac{1}{3} a_0 + \frac{1}{5} a_2 + \frac{1}{7} a_4 - J_2 = 0 & \frac{1}{5} a_1 + \frac{1}{7} a_3 - J_3 = 0 \\
(19) & \frac{1}{5} a_0 + \frac{1}{7} a_2 + \frac{1}{9} a_4 - J_4 = 0 & - J_1 a_1 - J_3 a_3 \qquad = C_2 \\
 & - J_0 a_0 - J_2 a_2 - J_4 a_4 \qquad = C_1 \, , &
\end{array}
$$

which are symmetric and which can be solved by the scheme of 23.3. In this case, either only the coefficients with even index or only those with odd index change when the degree of the approximating function is raised by one. If n is the degree of the approximating function, then for $n = 2m$ and $n = 2m + 1$, the coefficients with even index remain the same, while for $n = 2m - 1$ and $n = 2m$, those with odd index remain the same. The values of the coefficients are tabulated on page 315.

Degree of the Approximating Function	Even Coefficients	Odd Coefficients	Square of the Errors
$n = 0$	$a_0 = J_0$ (for $n = 0, 1$)		$C_0 = J_0^2$
$n = 1$		$a_1 = 3J_1$ (for $n = 1, 2$)	$C_1 = J_0^2 + 3J_1^2$
$n = 2$	$a_0 = \frac{3}{4}(3J_0 - 5J_2)$ (for $n = 2, 3$)		$C_2 = J_0^2 + 3J_1^2 + \frac{5}{4}(3J_2 - J_0)^2$
$n = 3$	$a_2 = \frac{15}{4}(3J_2 - J_0)$	$a_1 = \frac{15}{4}(5J_1 - 7J_3)$ (for $n = 3, 4$)	$C_3 = J_0^2 + 3J_1^2 + \frac{5}{4}(3J_2 - J_0)^2 + \frac{7}{4}(5J_3 - 3J_1)^2$
$n = 4$	$a_0 = \frac{15}{64}(15J_0 - 70J_2 + 63J_4)$ (for $n = 4, 5$)	$a_3 = \frac{35}{4}(5J_3 - 3J_1)$	$C_4 = J_0^2 + 3J_1^2 + \frac{5}{4}(3J_2 - J_0)^2 + \frac{7}{4}(5J_3 - 3J_1)^2 + \frac{9}{64}(35J_4 - 30J_2 + 3J_0)^2$
	$a_2 = \frac{105}{32}(-5J_0 + 42J_2 - 45J_4)$		
$n = 5$	$a_4 = \frac{315}{64}(4J_0 - 30J_2 + 35J_4)$	$a_1 = \frac{105}{64}(35J_1 - 126J_3 + 99J_5)$ $a_3 = \frac{315}{32}(-21J_1 + 90J_3 - 77J_5)$ $a_5 = \frac{693}{64}(15J_1 - 70J_3 + 63J_5)$	$C_5 = J_0^2 + 3J_1^2 + \frac{5}{4}(3J_2 - J_0)^2 + \frac{7}{4}(5J_3 - 3J_1)^2 + \frac{9}{64}(3J_0 - 30J_2 + 35J_4)^2 + \frac{11}{64}(63J_5 - 70J_3 + 15J_1)^2$

7. *Example*: We consider the example already treated in 8.4. The problem is to approximate $y = \sin x$ *in the interval* 0 to $\pi/2$ by a rational integral function of fourth degree. To change the limits to the values -1 and $+1$, we introduce a new variable t by the substitution

$$y = \sin\frac{\pi}{4}(t+1). \tag{19a}$$

Then

$$2J_0 = \int_{-1}^{+1} \sin\frac{\pi}{4}(t+1)\,dt = \frac{4}{\pi},$$

$$2J_1 = \int_{-1}^{+1} t\sin\frac{\pi}{4}(t+1)\,dt = -\frac{4}{\pi} + \left(\frac{4}{\pi}\right)^2$$

$$= -\frac{4}{\pi}(1 - 2J_0),$$

$$2J_2 = \int_{-1}^{+1} t^2\sin\frac{\pi}{4}(t+1)\,dt = \frac{4}{\pi} + 2\left(\frac{4}{\pi}\right)^2 - 2\left(\frac{4}{\pi}\right)^3$$

$$= \frac{4}{\pi}(1 - 2\cdot 2J_1), \tag{19b}$$

$$2J_3 = \int_{-1}^{+1} t^3\sin\frac{\pi}{4}(t+1)\,dt = -\frac{4}{\pi} + 3\left(\frac{4}{\pi}\right)^2 + 2\cdot 3\left(\frac{4}{\pi}\right)^3$$

$$- 2\cdot 3\left(\frac{4}{\pi}\right)^4 = -\frac{4}{\pi}(1 - 3\cdot 2J_2),$$

$$2J_4 = \int_{-1}^{+1} t^4\sin\frac{\pi}{4}(t+1)\,dt = \frac{4}{\pi} + 4\left(\frac{4}{\pi}\right)^2 - 3\cdot 4\left(\frac{4}{\pi}\right)^3$$

$$- 2\cdot 3\cdot 4\left(\frac{4}{\pi}\right)^4 + 2\cdot 3\cdot 4\left(\frac{4}{\pi}\right)^5 = \frac{4}{\pi}(1 - 4\cdot 2J_3).$$

From these we find as the approximating function, using the formulas shown in the table,

$$\bar{y} = 0.707102 + 0.554949t - 0.217987t^2 - 0.055165t^3 + 0.010901t^4. \tag{19c}$$

Also, we get

(19d) $$C_4 = 0.499999997.$$

Since $1/2 \int_{-1}^{+1} \sin^2 \pi/(t + 1)4 \, dt = 0.5$, then $M_0/2 = 3 \times 10^{-9}$ so that we get for the mean error of a value $m \approx 6 \times 10^{-5}$.

If we calculate 21 function values for $t = -1, 0.9, \cdots, 0, \cdots, +0.9, +1$, and plot the deviations from the true values, then we get

FIG. 108

the curve drawn in Fig. 108 as a solid line, while the dotted curve gives the errors for the approximating function obtained in 8.4 by the Newton interpolation formula. This latter curve has already been reproduced in Fig. 28 (page 87). If we form the mean error from the 21 calculated values, drawn according to 25(4), then we get $m \approx 8.6 \times 10^{-5}$, in sufficient agreement with the above value, while from the 19 values calculated by Newton's formula, we get $m_N \approx 22 \times 10^{-5}$.

8. The disadvantage of the approximation by a power series is that we must repeat the entire calculation whenever it turns out that the degree of the approximating function is chosen too low, because in addition to the further terms, the coefficients of the lower powers also change. We must therefore try to use, not simply a power series, but a rational integral function of mth degree such that the terms previously calculated no longer change when other terms are added. This is the case if only one coefficient appears each time in the normal equations, if, for example, only the terms appearing in the diagonal positions are different from zero. That is,

(20) $$\int_{-1}^{+1} g_\kappa(t) g_l(t) \, dt = \begin{cases} 0 \text{ for } \kappa \neq l \\ b_n \text{ for } \kappa = l. \end{cases}$$

In this case the normal equations become

(21) $$a_0 = \frac{1}{b_0} \int_{-1}^{+1} y \, dt, \quad a_1 = \frac{1}{b_1} \int_{-1}^{+1} y \cdot t \, dt, \cdots a_\kappa = \frac{1}{b_\kappa} \int_{-1}^{+1} y t^\kappa \, dt \cdots.$$

Functions which have this property are called orthogonal functions. The sum of the squares of the errors is also simplified in this case. It becomes

$$
(22)\qquad
\begin{aligned}
M_0 &= \int_{-1}^{+1} y^2\,dt - \int_{-1}^{+1} (a_0g_0(t) + a_1g_1(t) + \cdots + a_ng_n(t))^2\,dt \\
&= \int_{-1}^{+1} y^2\,dt - (b_0a_0^2 + b_1a_1^2 + b_2a_2^2 + \cdots + b_na_n^2).
\end{aligned}
$$

The approximating function which we get by a development in orthogonal functions is also a rational integral function. If we arrange this development according to increasing powers of x, the same approximating function is obtained which we had in the development by powers mentioned above.

9. The equation

$$
(23)\qquad \int_{-1}^{+1} g_\kappa(t)g_l(t)\,dt = \begin{cases} 0 & \text{for } \kappa \neq l \\ b_\kappa & \text{for } \kappa = l \end{cases}
$$

expresses a well-known property of the spherical harmonics which were introduced in 16.10:

$$
(24)\qquad P_\kappa(t) = \frac{1}{2^\kappa \cdot \kappa!}\frac{d^\kappa(t^2-1)^\kappa}{dx^\kappa}.
$$

For $l \geqq \kappa$ we find, by integration by parts,

$$
(25)\qquad
\begin{aligned}
\int_{-1}^{+1} P_\kappa(t)\cdot P_l(t)\,dt &= \frac{1}{2^{\kappa+l}\kappa!l!}\int_{-1}^{+1}\frac{d^\kappa(t^2-1)^\kappa}{dt^\kappa}\frac{d^l(t^2-1)^l}{dt^l}\,dt \\
&= \frac{(-1)^\kappa}{2^{\kappa+l}\kappa!l!}\int_{-1}^{+1}\frac{d^{2\kappa}(t^2-1)^\kappa}{dt^{2\kappa}}\frac{d^{l-\kappa}(t^2-1)^l}{dt^{l-\kappa}}\,dt \\
&= \frac{(-1)^\kappa(2\kappa)!}{2^{\kappa+l}\kappa!l!}\int_{-1}^{+1}\frac{d^{l-\kappa}(t^2-1)^l}{dt^{l-\kappa}}\,dt.
\end{aligned}
$$

If $l > \kappa$, then the integral contains the factor $t^2 - 1$ and is therefore zero at either limit; consequently it is shown that for the spherical harmonics

$$
(26)\qquad \int_{-1}^{+1} P_\kappa(t)\cdot P_l(t)\,dt = 0 \text{ for } \kappa \neq l.
$$

If $\kappa = l$, then we get from the above equation

$$
(26a)\qquad \int_{-1}^{+1} P_\kappa(t)^2\,dt = \frac{(-1)^\kappa(2\kappa)!}{2^{2\kappa}(\kappa!)^2}\int_{-1}^{+1}(t^2-1)^\kappa\,dt.
$$

The integral can be transformed still further by κ-fold application of integration by parts:

$$(27)\qquad \int_{-1}^{+1} (t^2 - 1)^\kappa \, dt = \frac{\kappa!(-1)^\kappa \cdot 2^\kappa}{1\cdot3\cdot5\cdot(2\kappa - 1)} \int_{-1}^{+1} t^{2\kappa} \, dt = \frac{(-1)^\kappa(\kappa!)^2 2^{2\kappa+1}}{(2\kappa + 1)!}.$$

If we substitute this, we find

$$(28)\qquad b_\kappa = \int_{-1}^{+1} P_\kappa(t)^2 \, dt = \frac{2}{2\kappa + 1}.$$

Consequently, we have found the coefficients of an approximation function of spherical harmonics,

$$(28a)\qquad \bar{y} = a_0P_0(x) + a_1P_1(x) + \cdots + a_nP_n(x).$$

If $y = f(t)$ is the function to be represented, this becomes

$$(29)\qquad a_\kappa = \frac{2\kappa + 1}{2} \int_{-1}^{+1} P(t)f(t) \, dt,$$

and the half sum of the squares of the errors becomes

$$(30)\qquad \frac{1}{2} M_0 = \frac{1}{2} \int_{-1}^{+1} f(t)^2 \, dt - \sum_0^n \frac{a_\kappa^2}{2\kappa + 1}.$$

The spherical functions of lowest order are

$$(31)\qquad \begin{aligned} P_0(x) &= 1 & P_1(x) &= x \\ P_2(x) &= \frac{3}{2}x^2 - \frac{1}{2} & P_3(x) &= \frac{5}{2}x^3 - \frac{3}{2}x \\ P_4(x) &= \frac{35}{8}x^4 - \frac{15}{4}x^2 + \frac{3}{8} & P_5(x) &= \frac{63}{8}x^5 - \frac{35}{4}x^3 + \frac{15}{8}x. \end{aligned}$$

10. The principal work is in the evaluation of the integral. In general this cannot be performed analytically; it is certainly not possible if we are dealing with functions which are empirical in origin. If these functions are given graphically, then we can obtain the quantities J_0, J_1, $\cdots$ in (16) by graphical integration. In this case, we must draw the curve for each integral, or we must use the integral curves of higher order. The equation of Jacobi [14(19)] becomes

$$(32)\qquad \begin{aligned} 2J_n = \int_{-1}^{+1} x^n y \, dx &= \int_{-1}^{+1} y \, dx - n \int_{-1}^{+1} \int_{-1}^{x} y \, dx^2 \\ &+ n(n - 1) \int_{-1}^{+1} \int_{-1}^{x} \int_{-1}^{x} y \, dx \cdots (-1)^n n! \int_{-1}^{+1} \int_{-1}^{x} \cdots \int_{-1}^{x} y \, dx^{n+1}, \end{aligned}$$

if we set $\xi = 0$, $a = -1$, $b = +1$. Then we can determine the desired integrals by the end ordinates of the multiple integral curve. But in

general, the method of drawing and measuring with a planimeter yields more accurate results.

If we do not want to carry out the calculation graphically, or if for some reason we do not want to use the methods given in the first section of this article, then we refer back to the calculation of the integrals by the approximation methods described in Chapter 3. If we have tables[3] for the spherical harmonics, we can simplify the work by calculation of the integrals $\int_{-1}^{+1} f(t)P_\kappa(t)\,dt$; otherwise we must first calculate the $\int_{-1}^{+1} t^\kappa f(t)\,dt$ individually; in this case the development in spherical functions has scarcely any advantages over the method of power series, in so far as computation is concerned. If the observation errors are small, then we can perhaps use Simpson's formula [15(21)] for equidistant ordinates, or formulas in which the ordinates enter in with as nearly equal weights as possible [15(5), (12), 16.7].

11. The expansions of this article can be extended to functions of several variables.[4] Let a function $z(u, v, w, \cdots)$ be given analytically, or in the form of tables with several entries, or graphically in the form of a nomogram, or, in the case of only two variables, in the form of a contour diagram. We seek an approximating function $\bar{z}$ such that

$$M = \int_\omega (z - \bar{z})^2\,do, \tag{33}$$

integrated over the whole region to be represented, is a minimum. If we have only discrete values, $[\epsilon\epsilon]$ must be a minimum. For example, with two variables,

$$\bar{z} = c_{00} + c_{10}u + c_{01}v + c_{20}u^2 + c_{11}uv + c_{02}v^2 \cdots . \tag{34}$$

Therefore

$$M = \int_\omega (z - c_{00} - c_{10}u - c_{01}v - c_{20}u^2 - c_{11}uv - c_{02}v^2 \cdots)^2\,do \tag{35}$$

must be made a minimum. This gives the normal equations

$$\begin{aligned}
&\int_\omega z(u, v)\,do - c_{00}\int_\omega do - c_{10}\int_\omega u\,do - c_{01}\int_\omega v\,do \\
&\qquad\qquad - c_{20}\int_\omega u^2\,do \cdots = 0, \\
&\int_\omega uz(u, v)\,do - c_{00}\int_\omega u\,do - c_{10}\int_\omega u^2\,do - c_{01}\int_\omega uv\,do
\end{aligned} \tag{36}$$

$$- c_{20} \int_\omega u^3 \, do \quad \cdots = 0,$$

$$\int_\omega vz(u, v) \, do - c_{00} \int_\omega v \, do - c_{10} \int_\omega uv \, do - c_{01} \int_\omega v^2 \, do$$

$$- c_{20} \int_\omega u^2 v \, do \cdots = 0.$$

. . . .

just as in the case of independent variables. If we combine terms in these normal equations, we obtain

$$\int_\omega z\bar{z} \, do - \int \bar{z}^2 \, do = 0, \tag{37}$$

so that here also we get

$$M_0 = \int_\omega z^2 \, do - \int_\omega z\bar{z} \, do = \int_\omega z^2 \, do - \int_\omega \bar{z}^2 \, do \tag{38}$$

for the minimum of M. In particular, upon introduction of rectangular coordinates, the normal equations become

$$\int_\omega z(x, y) \, dx \, dy - c_{00} \int_\omega dx \, dy - c_{10} \int_\omega x \, dx \, dy$$

$$- c_{01} \int_\omega y \, dx \, dy - c_{20} \int_\omega x^2 \, dx \, dy \cdots = 0,$$

$$\int_\omega xz(x, y) \, dx \, dy - c_{00} \int_\omega x \, dx \, dy - c_{10} \int_\omega x \, dx \, dy$$

$$- c_{01} \int_\omega xy \, dx \, dy - c_{20} \int_\omega x^3 \, dx \, dy \cdots = 0, \tag{39}$$

$$\int_\omega yz(x, y) \, dx \, dy - c_{00} \int_\omega y \, dx \, dy - c_{10} \int_\omega xy \, dx \, dy$$

$$- c_{01} \int_\omega y^2 \, dx \, dy - c_{20} \int_\omega x^2 y \, dx \, dy \cdots = 0.$$

If the function $z(x, y)$ is given in the form of an alignment chart, then

we can evaluate the above integrals with the aid of a planimeter. We can arrive at this goal more rapidly by using a moment planimeter because, with one circuit of the region, this gives

$$(40)\quad \int_\omega dx\,dy,\ \int_\omega x\,dx\,dy,\ \int_\omega x^2\,dx\,dy \text{ or } \int_\omega dx\,dy,\ \int_\omega y\,dx\,dy,\ \int_\omega y^2\,dx\,dy$$

for the corresponding choice of the curve. Therefore several coefficients of the normal equation are equal. If we draw the boundary curve of the region, $f_0(x, y)$, pointwise on a graph with two quadratic scales $\xi = x^2/2$, $\eta = y^2/2$, as is explained in 17.7b, then we get five coefficients of the normal equation by a double circuit of this curve with the moment planimeter. To get the integrals $\int_\omega z(x, y)\,dx\,dy \cdots$ (occurring as first terms in (39)) by measuring the contour lines with a planimeter, we must carry out a cubature, a plotting of these scale numbers as a function of the height, and repeated measurement (with a planimeter) of the surface results, just as is shown in 17.7a.

12. Numerically, this case is simplest if we deal with the representation of the path of the function in a rectangular region. We shall then so transform the variables that the bounding lines correspond to the values ± 1 of the variables, since then the normal equations are greatly simplified; for example, for the approximation by a function of third degree

$$(41)\quad \begin{aligned}\bar{z} = c_{00} + c_{10}x + c_{01}y + c_{20}x^2 + c_{11}xy + c_{02}y^2 + c_{30}x^3 + c_{21}x^2y \\ + c_{12}xy^2 + c_{03}y^3\end{aligned}$$

we get the ten normal equations,

$$(42)\quad \begin{aligned}
\frac{1}{4}\int z\,dx\,dy &= +c_{00} + \tfrac{1}{3}c_{20} + \tfrac{1}{3}c_{02} = J_{00}\\
\frac{1}{4}\int zx\,dx\,dy &= +\tfrac{1}{3}c_{10} + \tfrac{1}{5}c_{30} + \tfrac{1}{9}c_{12} = J_{10}\\
\frac{1}{4}\int zy\,dx\,dy &= +\tfrac{1}{3}c_{01} + \tfrac{1}{9}c_{21} + \tfrac{1}{5}c_{03} = J_{01}\\
\frac{1}{4}\int zx^2\,dx\,dy &= +\tfrac{1}{3}c_{00} + \tfrac{1}{5}c_{20} + \tfrac{1}{9}c_{02} = J_{20}\\
\frac{1}{4}\int xzy\,dx\,dy &= +\tfrac{1}{9}c_{11} = J_{11}
\end{aligned}$$

$$\frac{1}{4}\int zy^2 dx\,dy = +\frac{1}{3}c_{00} \quad +\frac{1}{9}c_{20} + \frac{1}{5}c_{02} \qquad = J_{02}$$

$$\frac{1}{4}\int zx^3\,dx\,dy = \quad +\frac{1}{5}c_{10} \qquad +\frac{1}{7}c_{30} \quad +\frac{1}{15}c_{12} \quad = J_{30}$$

$$\frac{1}{4}\int zx^2y\,dx\,dy = \quad +\frac{1}{9}c_{01} \qquad +\frac{1}{15}c_{21} \quad +\frac{1}{15}c_{03} = J_{21}$$

$$\frac{1}{4}\int zxy^2\,dx\,dy = \quad +\frac{1}{9}c_{10} \qquad +\frac{1}{15}c_{30} \quad +\frac{1}{15}c_{12} \quad = J_{12}$$

$$\frac{1}{4}\int zy^3\,dx\,dy = \quad +\frac{1}{5}c_{01} \qquad +\frac{1}{15}c_{21} \quad +\frac{1}{7}c_{03} = J_{03}\,.$$

From the fifth equation we get c_{11}, while the others can be collected in three groups of three equations, from each of which we can calculate three coefficients. Therefore we have

$$
\begin{aligned}
&c_{00} = \frac{7}{2}J_{00} - \frac{15}{4}(J_{20} + J_{02}), && c_{21} = \frac{135}{4}\left(J_{21} - \frac{1}{3}J_{01}\right),\\
&c_{20} = \frac{45}{4}\left(J_{20} - \frac{1}{3}J_{00}\right), && c_{10} = \frac{45}{2}J_{10} - \frac{105}{4}J_{30} - \frac{45}{4}J_{12}\,.\\
(43)\quad &c_{02} = \frac{45}{4}\left(J_{02} - \frac{1}{3}J_{00}\right), && c_{30} = \frac{175}{4}\left(J_{30} - \frac{3}{5}J_{10}\right),\\
&c_{01} = \frac{45}{2}J_{01} - \frac{105}{4}J_{03} - \frac{45}{4}J_{21}\,, && c_{12} = \frac{135}{4}\left(J_{12} - \frac{1}{3}J_{10}\right),\\
&c_{03} = \frac{175}{4}\left(J_{03} - \frac{3}{5}J_{01}\right). && c_{11} = 9J_{11}\,.
\end{aligned}
$$

If equidistant numerical values are given along the X axis as well as along the Y axis, we carry out the evaluation of the integrals $J_{00}, \cdots, J_{03}$ by the formula given in 13.9. This corresponds to the generalized Stirling formula. The theorem in Sec. 10 on the development in spherical harmonics of one variable is also valid for a development in Laplace's spherical harmonics of two variables. Because of the orthogonality of these functions, we are spared the solution of the equations (42). In breaking off with the terms of a certain order, we always have the best approximation in the sense of least squares. This hint will suffice here.

13. *Example*: In 13.7, the sagging of a wire, which depends on the length of the span s and the temperature t, was represented by a function of third degree, obtained from one of the interpolation formulas. A better approximation may be obtained if we calculate the coefficients according to the formulas given in Sec. 12. For this purpose, we first transform the limits to the values ± 1 by the transformation

$$x = \frac{s - 140}{100}, \qquad y = \frac{t - 10}{30}, \tag{43a}$$

and then form the approximation values for the integrals by means of the formula 13.9. This gives

$$\begin{aligned} J_{00} &= 431.544, & J_{10} &= 189.502, & J_{20} &= 163.396, & J_{30} &= 112.841, \\ J_{01} &= 13.270, & J_{11} &= 1.406, & J_{21} &= 3.791, \\ J_{02} &= 143.769, & J_{12} &= 62.856, \\ J_{03} &= 8.047. \end{aligned} \tag{43b}$$

From these we can get the coefficients of the approximation formula:

$$\begin{aligned} c_{00} &= 358.535, & c_{10} &= 594.589, & c_{20} &= 219.915, & c_{30} &= -37.634, \\ c_{01} &= 44.693, & c_{11} &= 12.654, & c_{21} &= -21.341, \\ c_{02} &= -0.889, & c_{12} &= -10.508, \\ c_{03} &= 3.719. \end{aligned} \tag{43c}$$

If we calculate the 77 given function values by the formulas of the previous section, we get the following errors

		40 m	60 m	80 m	100 m	120 m	140 m	160 m	180 m	200 m	220 m	240
		−1	−0.8	−0.6	−0.4	−0.2	0	+0.2	+0.4	+0.6	+0.8	+
−20°	−1	+4.7	−3.2	−1.6	+5.7	−2.1	−0.8	−1.1	−3.0	−13.2	−8.5	+25
−10°	$-\frac{2}{3}$	+3.5	−1.8	−0.2	+6.5	−1.4	−1.8	−0.4	+2.9	− 7.7	−3.0	+10
0°	$-\frac{1}{3}$	+2.9	−1.4	−2.3	+4.5	−5.1	−1.6	+1.0	+4.0	− 4.5	+0.8	+17
+10°	0	+1.5	−2.1	−3.9	+3.3	−3.3	−1.5	0	+5.2	− 3.7	−0.3	+15
+20°	$+\frac{1}{3}$	+2.3	−2.2	−6.4	+1.9	−3.2	−0.5	+1.2	+4.2	+ 0.5	−1.5	+11
+30°	$+\frac{2}{3}$	+5.0	−0.7	−5.8	0	−4.1	+1.0	+2.6	−0.1	− 1	−4.9	+ 5
+40°	1	+8.5	+5.0	−2.4	+6.5	+3.0	+3.1	+3	−0.9	0	−8.7	− 1

As the sum of the squares of the errors for these 77 values, we get $M_0 = 2553$, while for the formula given in 13(7), we get $M = 4243$.

Since 10 coefficients are calculated here, we have as the mean error of a single measurement

$$m_0 = \left(\frac{2533}{67}\right)^{1/2} \approx 6.2 \qquad \text{or} \qquad m = \left(\frac{4243}{67}\right)^{1/2} \approx 8.$$

If we are interested in the largest percent accuracy possible, then the formula which makes the sum of the squares of the absolute errors a minimum is not necessarily the best. In such a case we must make the sum of the squares of the relative errors a minimum, rather than the sum of the squares of the absolute errors.[5]

An even better representation would be obtained if we were to apply the methods of Sec. 1 to 6 for discrete values, since then the inaccuracy in the calculation of the integral is avoided.

NOTES

1. The smoothing out of data is handled in detail in Blaschke, *Vorlesungen über athematische Statistik* (Leipzig, 1906), Ch. VI: also in Whittaker and Robinson, *The alculus of Observations*, 2nd ed. (London, 1926), Ch. IX.

2. Cassebaum, *Über das Verhalten von weichem Flussstahl jenseits der Proportional-ätsgrenze.* Dissertation (Gottingen, 1910), p. 59.

3. For example, Jahnke-Emde, *Funktionstafeln* (Leipzig, 1909), p. 83 ff.

. Cf. *Enzyklopädie d. math. Wiss.* IV, 1, 10; Schmidt, *Erdmagnetismus* No. 20, and the literature listed there.

5. This type of representation could also be used, for example, in the investigation of the density distribution in star clusters. E.g., *Naturwissenschaften* XV (1927), p. 243.

28. Approximation of the Entire Course of Periodic Functions.

1. With periodic functions we can be satisfied with the *approximate epresentation of one period* if we make use of such periodic functions for he approximation whose period coincides with the period taken as the nown period, or if we use a period which is a submultiple of the period f the given function. We choose the sine and cosine as the simplest functions, which are everywhere finite and periodic. If the length of the period f the given function is d, then we transform this to the legnth 2π by a hange in the argument $t = 2\pi x/d$, and use as the approximation a sum f sine and cosine functions of periods

$$2\pi, \pi, \frac{2\pi}{3}, \frac{\pi}{2}, \frac{2\pi}{5}, \frac{\pi}{3}, \cdots$$

Therefore, we set

$$\begin{aligned}\bar{y} &= a_0 + a_1 \cos t + a_2 \cos 2t + a_3 \cos 3t \cdots \\ &\quad + b_1 \sin t + b_2 \sin 2t + b_3 \sin 3t \cdots .\end{aligned} \tag{1}$$

When we consider the entire path of the given function $y = f(t)$, the normal equations 25(12) take the form

$$\int_0^{2\pi} (y - \bar{y})\,dt = 0, \qquad \int_0^{2\pi} (y - \bar{y}) \cos lt\,dt = 0, \tag{2}$$
$$\int_0^{2\pi} (y - \bar{y}) \sin lt\,dt = 0.$$

Now the functions used for the approximation here form a system of orthogonal functions. Then

$$\int_0^{2\pi} dt = 2\pi, \qquad \int_0^{2\pi} \cos lt\,dt = 0, \qquad \int_0^{2\pi} \sin lt\,dt = 0.$$

$$\int_0^{2\pi} \sin lt \sin \kappa t\,dt = \frac{1}{2}\int_0^{2\pi} \cos(l - \kappa)t\,dt = \frac{1}{2}\int_0^{2\pi} \cos(l + \kappa)t\,dt$$
$$= \begin{cases} 0 \text{ for } \kappa \neq l \\ \pi \text{ for } \kappa = l \end{cases}$$

(3)

$$\int_0^{2\pi} \sin lt \cos \kappa t\,dt = \frac{1}{2}\int_0^{2\pi} \sin(l + \kappa)t\,dt + \frac{1}{2}\int_0^{2\pi} \sin(l - \kappa)t\,dt = 0$$

$$\int_0^{2\pi} \cos lt \cos \kappa t\,dt = \frac{1}{2}\int_0^{2\pi} \cos(l - \kappa)t\,dt + \frac{1}{2}\int_0^{2\pi} \cos(l + \kappa)t\,dt$$
$$= \begin{cases} 0 \text{ for } \kappa \neq l \\ \pi \text{ for } \kappa = l. \end{cases}$$

In the normal equations 25(12) therefore, all terms except those on the main diagonal become zero, so that we can write the coefficients without further observation. They are

$$a_0 = \frac{1}{2\pi}\int_0^{2\pi} f(t)\,dt, \quad a_l = \frac{1}{\pi}\int_0^{2\pi} f(t) \cos lt\,dt, \tag{4}$$
$$l = 1, 2, \cdots, n$$
$$b_l = \frac{1}{\pi}\int_0^{2\pi} f(t) \sin lt\,dt.$$

All integrals of the diagonals terms are π except the first, which has the value 2π. If we again introduce the period d by the substitution $t = 2\pi x/d$, we have

$$a_0 = \frac{1}{d}\int_0^d f(x)\,dx, \qquad a_l = \frac{2}{d}\int_0^d f(x)\cos l\frac{2\pi}{d}x\,dx,$$

(5)

$$b_l = \frac{2}{d}\int_0^d f(x)\sin l\frac{2\pi}{d}x\,dx.$$

If we judge the value of the approximation by the value of the integral

(6) $$M = \int_0^{2\pi} (f(t) - \bar{y})^2\,dt,$$

this representation gives the best approximation value at the term at which we break it off, since when further terms are added, the coefficients which had previously been calculated do not change.

2. The expression (6) can be transformed into

(7) $$M_0 = \int_0^{2\pi} y^2\,dt - \int_0^{2\pi} y\bar{y}\,dt = \int_0^{2\pi} y^2\,dt - \int_0^{2\pi} \bar{y}^2\,dt$$

by the use of 25(18). If we carry out the quadrature in the last term and consider the equations (3), then, because of the orthogonality, we have

(8) $$M_0 = \int_0^{2\pi} y^2\,dt - \pi(2a_0^2 + a_1^2 + a_2^2 \cdots a_n^2 + b_1^2 + b_2^2 + \cdots + b_n^2).$$

We can again take the root of the mean square deviation

(9) $$m = \left(\frac{M_0}{2\pi}\right)^{1/2} = \left(\frac{1}{2\pi}\int_0^{2\pi} y^2\,dt - a_0^2 - \frac{1}{2}\sum_{l=1}^{l=n}(a_l^2 + b_l^2)\right)^{1/2}$$

as *a measure of the accuracy obtained.*

3. In many cases it is desirable *to join the two terms with the same period.* If in

(10) $$a_l \cos lt + b_l \sin lt$$

we substitute

(11) $$a_l = r_l \sin \varphi_l, \qquad b_l = r_l \cos \varphi_l,$$

i.e.,

(12) $$r = (a_l^2 + b_l^2)^{1/2}, \qquad \operatorname{tg}\varphi_l = \frac{a_l}{b_l},$$

then the two terms become

(13) $$r_l \sin(lt + \varphi_l).$$

If we choose the phase angle φ_l (on consideration of the signs of a_l and b_l) in the right quadrant, then the amplitude of the sine wave is positive. Therefore we can express the periodic function as a sum of sine terms with various phase angles. The mean error then becomes

$$(14) \qquad m = \left(\frac{1}{2\pi}\int_0^{2\pi} y^2\,dt - a_0^2 - \frac{1}{2}\sum_{l=1}^{l=n} r_l^2\right)^{1/2}.$$

4. The *calculation of the coefficients* $a_0, \cdots, a_n$ and $b_1, \cdots, b_n$ must be carried out by approximation methods for functions which are given empirically. If we want to consider the entire path of some curve given by an oscillogram, then we need consider only graphical or instrumental methods. A graphical method has been given by von Mises.[1] The coefficient a_0, which is the mean ordinate of the entire curve, can be determined by ordinary graphical integration (Art. 14). The coefficients a_l and b_0 (4), which we write in the form

$$(15) \qquad a_l = \frac{1}{l\pi}\int_0^{2\pi} f(t)d(\sin lt); \qquad b_l = -\frac{1}{l\pi}\int_0^{2\pi} f(t)d(\cos lt)$$

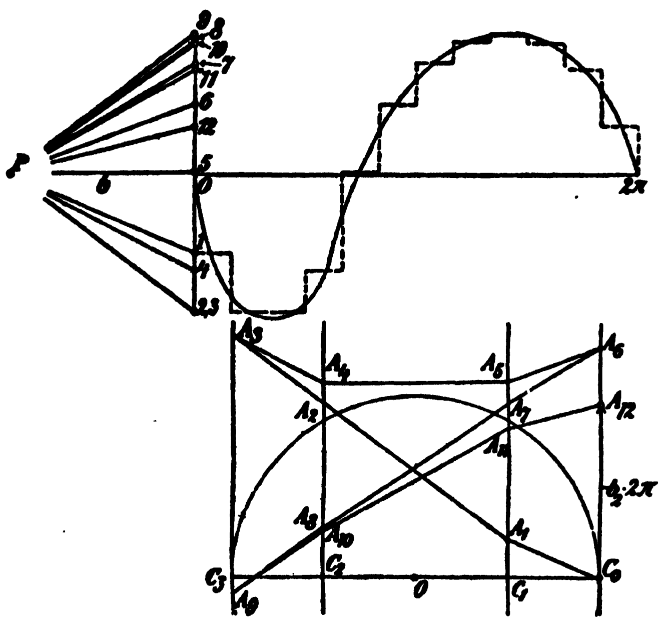


FIG. 109

would necessitate a drawing before the integration. We must shift tl

ordinates on the abscissa t so that they get the abscissas sin lt or cos lt. In this way, between the y parallels at distances $+1$ and -1, we would get l ascending and descending curves. We avoid the details of drawing in the following way: we divide the entire interval into q equal parts and construct in each strip the mean ordinate, in a way still to be mentioned. This determines an appropriate direction in a pencil or rays. The slope of this ray is tg $\varphi_m = y_m/b$ where b is the integration base. A circle with radius r is also divided into q equal parts (in Fig. 109, $q = 12$), and parallels to the y axis are drawn through the division points. This is for the construction of the coefficients. Now we draw a parallel to the first direction ray $P1$, beginning from the intersection point C_0 of the circle with the x axis. This parallel cuts the y parallel through the circle division point (which is $rl(2\pi)/q$ radians from C_0, in Fig. 109, $l = 2$, $r = 2.5$ cm., $b =$ 2.5 cm.) in A_1. Then

$$\text{(15a)} \quad C_1A_1 = \operatorname{tg} \varphi_1 r\left(\cos \frac{l2\pi}{q} - \cos 0\right) = \frac{r}{b} y_1\left(\cos \frac{l2\pi}{q} - \cos 0\right).$$

If we draw through A_1 a parallel to the second direction ray $P2$ up to the intersection A_2 with the y-parallel through the point with angular measure $r \cdot 2l \cdot 2\pi/q$, then

$$\text{(15b)} \qquad C_2A_2 = C_1A_1 + \frac{r}{b} y_2\left(\cos 2l \frac{2\pi}{q} - \cos l \frac{2\pi}{q}\right) \text{ etc.}$$

The last direction parallel cuts off a length

$$\text{(15c)} \qquad \begin{aligned} C_0A_{12} &= \frac{r}{b} \sum_{m=1}^{q} y_m\left(\cos ml \frac{2\pi}{q} - \cos (m-1)l \frac{2\pi}{q}\right) \\ &= \frac{r}{b} \sum_{m=1}^{q} y_m \Delta\left(\cos ml \frac{2\pi}{q}\right), \end{aligned}$$

a value which agrees, for correct choice of the mean ordinate y_m, with the integral $\int_0^{2\pi} f(t)d$ (cos lt), except for the factor r/b. The value y_m would be the mean ordinate, in the mth interval, of the curve drawn on the abscissa cos lt. This could be found by the rule given in Art. 14. For example, for the approximation by parabolas (with axes parallel to the y axis) this could be done by dividing the part of the mean parallel between chord and curve into three parts. By dividing the length C_mC_{m-1} on the circle diameter, we can now find the value t_m from the corresponding arc length belonging to this mean parallel. By estimation with the eye, we can determine the ratio in which this point divides the arc between the points $r \cdot ml \cdot 2\pi/q$ and $r(m - 1)l \cdot 2\pi/q$ in the corresponding abscissa interval. We find the mean ordinate of the division interval for the corresponding integral curve by dividing into three parts the length between

the corresponding point on the curve and the mean position on the chord (Fig. 110). In general it is only necessary to choose the mean ordinate in a few intervals for each integration.

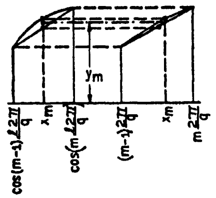


FIG. 110

The construction of the values a_1 to a_n follows in a completely similar fashion; only instead of the y-parallels, we employ parallels to the t axis, and in place of the direction rays, lines perpendicular to these. Also, we begin the integration polygon at the center of the circle.

The construction of the coefficients $a_0, \cdots, a_3, b_1, \cdots, b_3$ is shown

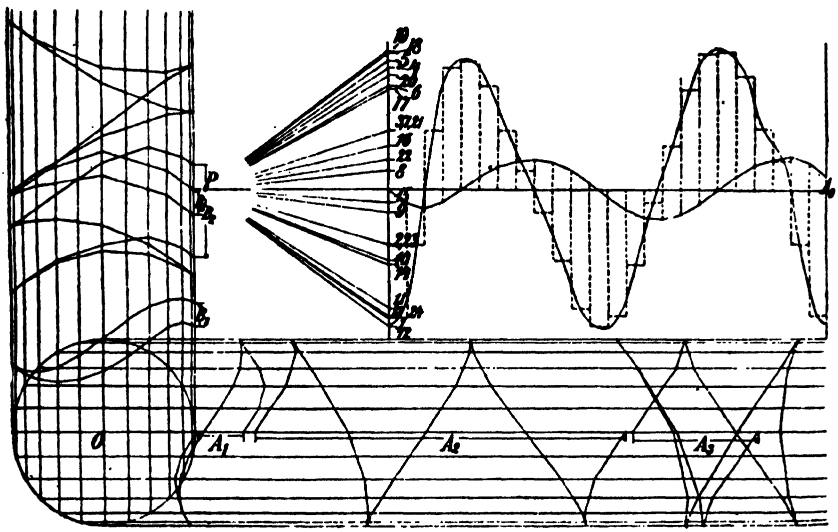


FIG. 111

in Fig. 111. In this case, $b = 37.5$ mm., and the scale modulus is $90/2\pi$ mm., $r = 18.75$ mm., so that

$$a_0 = \frac{37.5 \times A_0}{90} \text{ mm}; \; a_l = \frac{2A_l}{l\pi} \text{ mm}; \; b_l = -\frac{2B_l}{l\pi} \text{ mm}. \; (l = 1, 2, 3),$$

if we measure A_0, A_l and B_l in millimeters. The constant a_0 is found by

ordinary integration, while the B corresponding to the above development are constructed on the left side of the figure, the A under the curve. While in the construction of the coefficients with index 1 we must change the direction of the integration polygon for each parallel, in those with the index 2, one parallel is calculated, in those with the index 3 two parallels, etc. If the endpoint lies underneath or to the left of the starting point of the integration polygon, then the coefficient to be calculated is positive, otherwise negative. We then find

$$a_0 = +1.17 \text{ mm.},$$

$$a_1 = +6.05 \text{ mm.}, \; a_2 = -24.15 \text{ mm.}, \; a_3 = -5.4 \text{ mm.}$$

$$b_1 = -2.95 \text{ mm.}, \; b_2 = +\;5.75 \text{ mm.}, \; b_3 = -1.06 \text{ mm.}$$

5. Another possibility for the determination of the coefficients of the Fourier series over the entire path of a curve is offered by the harmonic *analyzer*. We shall describe the analyzer of *Mader*. The constant a_0 of the series can be found with a planimeter. The coefficients a_l and b_l can be put into the form

$$(16) \quad a_l = \frac{1}{l\pi}\int_0^d f(x)d\left(\sin l\,\frac{2\pi x}{d}\right); \qquad b_l = -\frac{1}{l\pi}\int_0^d f(x)d\left(\cos l\,\frac{2\pi x}{d}\right)$$

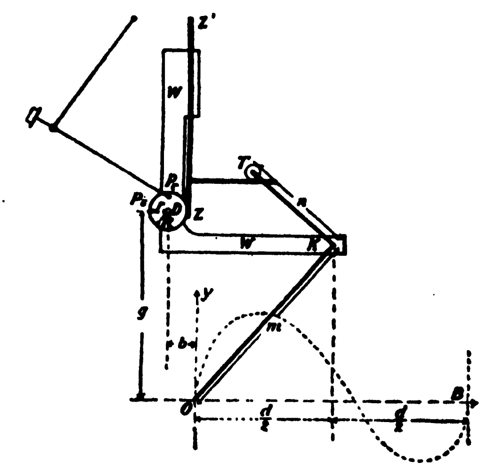


FIG. 112

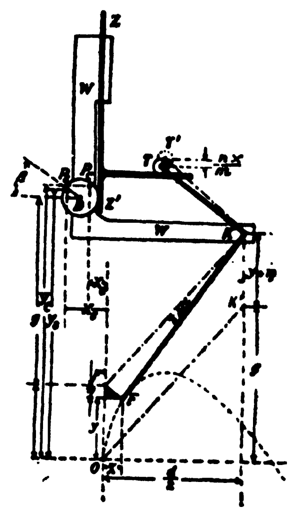


FIG. 113

if we start from the equations (5). This is done in order to handle arbitrary periods. The Mader analyzer evaluates this form.

It consists of a carriage WW which can be shifted parallel to the y axis

by means of a milled wheel running in the groove of a rail not shown in Fig. 112. This carriage possesses three movable parts:

a. The first part is a right angled lever OKT, whose axis K moves on the y-parallel through the middle of the period, i.e., at the distance $d/2$ from the y axis. By means of the roller T this moves

b. a rack ZZ', which can also be shifted on the carriage only in the direction of the y axis, and which has an arm parallel to the x axis, on which T moves. This rack turns

c. a disk of radius R with two insertion points P_s and P_c at a distance r from the center D. The tracing point of a planimeter can be set in these holes.

If the tracing point F is at the starting point 0, then the coordinates are

$$\begin{aligned} P_s &: x_s = -(b + r), \qquad & y_s &= g, \\ P_c &: x_c = -b, \qquad & y_c &= g + r. \end{aligned} \tag{17}$$

If now we trace along with the point F on the curve $y = f(x)$, we can consider this motion as consisting of the following two motions:

a. Displacement of the carriage WW a distance $y + \eta$ parallel to the y axis, the lever $F'K'T'$ remaining fixed;

b. Rotation of the lever from $F'K'T'$ to FKT through the angle $\Delta\alpha$. If the arms of the lever have the lengths m and n, then, by means of the roller T, this causes a displacement of the rack ZZ' by nx/m, as we can see from the similarity of the small black triangles. This displacement gives rise to a rotation of the disk D through the angle $\beta = nx/mR$. The new coordinates of P_s and P_c are then

$$\begin{aligned} P_s &: x_s = -(b + r\cos\beta), \qquad & y_s &= g + y + \eta + r\sin\beta, \\ P_c &: x_c = -b + r\sin\beta, \qquad & y_c &= g + y + \eta + r\cos\beta. \end{aligned} \tag{18}$$

If we now traverse the curve $y = f(x)$ with the point F, and then trace back on the axis to the starting point, then the area of the curve described by P_s or P_c (which is measured by the planimeter) becomes

$$\begin{aligned} J_s &= +\int_0^d (g + y + \eta + r\sin\beta)d(-b - r\cos\beta) \\ &\quad + \int_d^0 (g + \eta + r\sin\beta)d(-b - r\cos\beta) \\ &= -r\int_0^d y d(\cos\beta) = -r\int_0^d y d\cos\frac{nx}{mR}, \end{aligned}$$

(19)

$$J_c = \int_0^d (g + y + \eta + r\cos\beta)d(-b + r\sin\beta)$$

$$+ \int_d^0 (g + \eta + r\cos\beta)d(-b + r\sin\beta)$$

$$= +r\int_0^d yd(\sin\beta) = r\int_0^d yd\left(\sin\frac{nx}{mR}\right).$$

We then make

$$R = d\frac{n}{m}\frac{1}{2l\pi}, \qquad r = \frac{\kappa}{l\pi} \qquad (l = 1, 2\cdots), \tag{20}$$

so that this becomes

(20a)
$$J_s = -\frac{\kappa}{l\pi}\int_0^d f(x)d\left(\cos l\frac{2\pi x}{d}\right) = \kappa b_l\,,$$
$$J_c = \frac{\kappa}{l\pi}\int_0^d f(x)d\left(\sin l\frac{2\pi x}{d}\right) = \kappa a_l\,.$$

Therefore, according to whether we put the planimeter point in the hole P_s or P_c, we get as the difference of the initial and final reading the κ fold coefficient of the sine or cosine term.

From (20) it follows that a special disk is necessary for each pair of coefficients. This disk is removable and for each pair of coefficients, one, or for $l = 7, 9, 11, \cdots$ two separate disks are inserted. The adjustment of the apparatus is made by means of a scale inserted in the groove of the rail, and running perpendicular to it. This scale has half millimeter divisions, running in both directions from the point inserted at the middle of the interval. The period can also be determined with this. The tracing point is adjustable on the angle arm FK, and is put in at the division mark on FK which gives the numerical measure of the period in millimeters. In this way it becomes possible to analyze curves of a period length from 2 to 36 cm., without any necessity of carrying out a drawing of the given curve with a particular period length. The constant is 10 with the apparatus used, so that, with the ordinary polar planimeter (1 revolution equals 1 sq. m.) one vernier unit corresponds to 0.1 mm. amplitude. If there is no gear slipping between the wheel D and the rack ZZ', the error of the apparatus is not very large.[2]

NOTES

1. Friesecke, *Z. f. angew. Math. u. Mech.* 2 (1922), p. 313.
2. Other analyzers are described in Willers, *Mathematische Instrumente* (Berlin, 1926), Art. 15, and Art. 16.

29. Harmonic Analysis of Periodic Functions Which Are Given Discrete Values.

1. Frequently only discrete values of the periodic function to be evaluated will be given. In the application of *numerical methods*, in particular, we shall be able to consider only discrete values, even if the entire path of the function is given. In the following it is now assumed that these values are equidistant. Suppose that the scale is already so chosen that the modulus is 2π. If we then divide the interval into q equal parts, so that corresponding to the $q + 1$ abscissa values

$$x_0 = 0, x_1 = \frac{2\pi}{q}, x_2 = 2\frac{2\pi}{q}, \cdots, x_m = m\frac{2\pi}{q}, \cdots, x_q = 2\pi,$$

we have the $q + 1$ function values $y_0, y_1, \cdots, y_q$, where in general, $y_0 = y_q$. Therefore we now have to determine the approximate $\bar{y}$ so that, by 25(2), $\sum_{m=1}^{m=q} (y_m - \bar{y}(x_m))^2$ becomes a minimum. In this method, only the best possible approximation of the discrete function values is important; we are not concerned with the rest of the path of the function. Under certain circumstances this can lead to very sharp deviations for the intermediate values, if these intervals are not made so narrow that large intermediate oscillations are avoided.

We take as the approximation function

$$(1)\qquad \begin{aligned}\bar{y} = a_0 + a_1 \cos x + a_2 \cos 2x + a_3 \cos 3x + \cdots + a_n \cos nx \\ + b_1 \sin x + b_2 \sin 2x + b_3 \sin 3x + \cdots + b_n \sin nx.\end{aligned}$$

The minimum condition, 25(10), leads to the normal equations

$$(2)\qquad \begin{aligned}&\sum_{m=1}^{m=q} (y_m - \bar{y}(x_m)) = 0, \\ &\sum_{m=1}^{m=q} (y_m - \bar{y}(x_m)) \cos lx_m = 0, \qquad m = 1, 2 \cdots q, \quad l = 1, 2 \cdots n. \\ &\sum_{m=1}^{m=q} (y_m - \bar{y}(x_m)) \sin lx_m = 0\end{aligned}$$

This gives $2n + 1$ equations from which the $2n + 1$ constants $a_0, \cdots, a_n, b_1, \cdots, b_n$ are to be determined. Now the entire problem has a meaning only if $2n + 1 \leqq q + 1$ because the $q + 1$ values $y_0, \cdots, y_q$ should approximate the function $\bar{y}$ as closely as possible. But if more than $q + 1$ constants appear in this function, then there are an infinite number of ways that $\bar{y}$ can take on these $q + 1$ values. The determination of the coefficients is then no longer unique. If $2n + 1 = q + 1$ the total error can be made equal to zero; i.e., $[y_m - y(x_m)]^2 = 0$. If $2n + 1 < q + 1$, then it is only possible to reduce this value to a minimum, not zero,

2. For the simplification of the normal equations, we must observe that in the case $l < q$

$$\sum_{m=1}^{m=q} \sin lx_m = 0 \quad \text{and} \quad \sum_{m=1}^{m=q} \cos lx_m = \begin{cases} q \\ 0, \end{cases} \tag{3}$$

according as l is zero or has another integer value. The equations (3) are easily derived if we join them together by use of the Euler equation

$$e^{i\varphi} = \cos \varphi + i \sin \varphi. \tag{3a}$$

This gives

$$\sum_{m=1}^{m=q} (\cos lx_m + i \sin lx_m) = \sum_{m=1}^{m=q} e^{ilx_m} = \sum_{m=1}^{m=q} e^{ilm(2\pi/q)}. \tag{3b}$$

If we now set $e^{il(2\pi/q)} = r$, then we get

$$\begin{aligned} \sum_{m=1}^{m=q} e^{ilm(2\pi/q)} &= r + r^2 + r^3 + \cdots + r^q = r\frac{r^q - 1}{r - 1} \\ &= e^{il(2\pi/q)} \frac{e^{il2\pi} - 1}{e^{il(2\pi/q)} - 1}. \end{aligned} \tag{3c}$$

For all cases in which $l < q$ is not zero, the numerator of the fraction of the right side is zero, while the denominator is not zero. Therefore the sum (and consequently its real and imaginary parts) is zero. But if $l = 0$, then $r = e^{il(2\pi/q)} = 1$ so that the real part of the sum is q, while the imaginary part is again zero.

If we now substitute in (2) the expression (1) for $\bar{y}$, then sums of the products of the trigonometric functions appear. These can be transformed by use of the equations (3). Let κ, λ, n and s be any integers, and let, say $\kappa > \lambda$, then

$$\sum_{m=1}^{m=q} \sin \kappa x_m \sin \lambda x_m = \frac{1}{2}\sum_{m=1}^{m=q} \cos(\kappa - \lambda)x_m - \frac{1}{2}\sum_{m=1}^{m=q} \cos(\kappa + \lambda)x_m$$

$$= \begin{cases} 0 \text{ for } \kappa \neq s\cdot q \pm \lambda \\[2ex] \frac{q}{2} \text{ for } \kappa = s\cdot q + \lambda \quad \lambda \neq n\cdot\frac{q}{2} \\[2ex] -\frac{q}{2} \text{ for } \kappa = s\cdot q - \lambda \quad \lambda \neq n\cdot\frac{q}{2} \\[2ex] 0 \text{ for } \kappa = s\cdot q \pm \lambda \quad \lambda = n\cdot\frac{q}{2}, \end{cases} \tag{4}$$

$$\sum_{m=1}^{m=q} \cos \kappa x_m \cos \lambda x_m = \frac{1}{2} \sum_{m=1}^{m=q} \cos(\kappa - \lambda)x_m + \frac{1}{2} \sum_{m=1}^{m=q} \cos(\kappa + \lambda)x_m$$

$$= \begin{cases} 0 \text{ for } \kappa \neq s \cdot q + \lambda \\ \frac{q}{2} \text{ for } \kappa = s \cdot q \pm \lambda \quad \lambda \neq n \cdot \frac{q}{2} \\ q \text{ for } \kappa = s \cdot q \pm \lambda \quad \lambda = n \cdot \frac{q}{2}, \end{cases}$$

$$\sum_{m=1}^{m=q} \sin \kappa x_m \cos \lambda x_m = \frac{1}{2} \sum_{m=1}^{m=q} \sin(\kappa + \lambda)x_m + \frac{1}{2} \sum_{m=1}^{m=q} \sin(\kappa - \lambda)x_m = 0.$$

If we substitute this in the normal equations, then these become, for $l < q$:

$$\text{(5)} \qquad \begin{aligned} &\sum_{m=1}^{m=q} y_m = q \cdot a_0 , \\ &\sum_{m=1}^{m=q} y_m \cos lx_m = \frac{q}{2} a_l , \qquad \sum_{m=1}^{m=q} y_m \sin lx_m = \frac{q}{2} b_l . \end{aligned}$$

3. In this way, if $n = q/2$, there is the additional equation

$$\text{(6)} \qquad \sum_{m=1}^{m=q} (-1)^m y_m = q \cdot a_{q/2} .$$

If we collect the equations (2) with $a_0 , \cdots , a_l , b_1 , \cdots , b_l$, then we get

$$\text{(6a)} \qquad \sum_{m=1}^{m=q} (y_m - \bar{y}(x_m))\bar{y}(x_m) = 0.$$

The expression for the sum of the squares of the errors is transformed by the use of this relation as in 25(16); it becomes

$$\text{(7)} \qquad M_0 = \sum_{m=1}^{m=q} (y_m - \bar{y}(x_m))^2 = \sum_{m=1}^{m=p} y_m^2 - \sum_{m=1}^{m=q} \bar{y}(x_m)^2,$$

and if we consider the equations (4), it follows, if $n < q/2$, that

$$\text{(8)} \qquad M_0 = \sum_{m=1}^{m=q} y_m^2 - \frac{1}{2} q(2a_0^2 + a_1^2 + b_1^2 + a_2^2 + b_2^2 \cdots a_n^2 + b_n^2).$$

If $n = q/2$, then this becomes

$$M_0 = \sum_{m=1}^{m=q} y_m^2 - \frac{1}{2}q(2a_0^2 + a_1^2 + b_1^2 + a_2^2 + b_2^2 + \cdots \tag{9}$$
$$+ a_{n-1}^2 + b_{n-1}^2 + 2a_n^2).$$

4. In each case it is practical to choose, as the *number of function values to be used for the calculation, a number divisible by* 4, i.e., to set $q = 4h$, since then the values of the sine and cosine are repeated in the four quadrants and we can combine certain function values at the outset. Furthermore, if we assume that the number of coefficients to be determined is equal to the number of function values used, then $M_0 = 0$, and the normal equations (7) and (8) become

$$\sum_{m=1}^{m=4h} y_m = a_0 \cdot 4h, \qquad \sum_{m=1}^{m=4h} (-1)^m y_m = a_{2h} \cdot 4h,$$

$$\left.\begin{aligned} \sum_{m=1}^{m=4h} y_m \cos lx_m &= a_l \cdot 2h \\ \sum_{m=1}^{m=4h} y_m \sin lx_m &= b_l \cdot 2h \end{aligned}\right\} \quad l = 1, 2, \cdots, 2h - 1. \tag{10}$$

Because of the great importance of harmonic analysis for applied work, a series of processes have been worked out to calculate these sums as mechanically, and with as little computation work as possible. First we shall describe a *method for the graphical determination of the coefficients.* If the equidistant ordinates are given as lengths, then we can obtain a_0 and a_{2h} simply by superposition of these lengths on a line in the direction determined by the sign. From the other equations we join the two terms with each index:

$$2h(a_l + ib_l) = \sum_{m=1}^{m=4h} y_m(\cos lx_m + i \sin lx_m) = \sum_{m=1}^{m=4h} y_m e^{ilx_m}. \tag{10a}$$

To lessen the graphical work, we observe that $e^{ilx_m} = (-1)^l e^{il(x_m+\pi)}$ so that if we form the ordinate sums and differences according to the scheme

$$\begin{array}{cccccc} y_1 & y_2 & y_3 & \cdots & y_{2h-1} & y_{2h} \\ y_{2h+1} & y_{2h+2} & y_{2h+3} & \cdots & y_{4h-1} & y_{4h} \\ \hline s_1 & s_2 & s_3 & \cdots & s_{2h-1} & s_{2h} \\ d_1 & d_2 & d_3 & \cdots & d_{2h-1} & d_{2h} \end{array}, \tag{10b}$$

we can also write the above sums

$$2h(a_l + ib_l) = \sum_{m=1}^{m=2h} d_m e^{ilx_m} \text{ for odd } l$$

$$2h(a_l + ib_l) = \sum_{m=1}^{m=2h} s_m e^{ilx_m} \text{ for even } l.$$

The d_m and s_m appear here as vectors which form the angles lx_m with the positive x axis. If we add these vectors, the vector sum is $2h(a_l + ib_l)$. For each value l a polygon is determined which gives the two coefficients with the index l as the components of the final vector. It is simplest to mark the directions on a circle about the origin, which form the angle $x_m = 2\pi m/q$ with one another. In the case of twelve ordinates, therefore $h = 3$, we get the directions of the polygon sides simply by a parallel shifting of a drawing triangle with the angles 30° and 60°. For a larger number, say 16 or 32, of equidistant ordinates, we can obtain the directions conveniently with the direction ruler of von Sanden.[1] Following a suggestion of Groeneveld,[2] we can also use special paper for the construction, which has parallel bands in the necessary directions. The individual lines of these bands are so close together that a strip of transparent coordinate paper, on which the lengths s_m and d_m are plotted additively (on a longitudinal line placed in the corresponding direction), always covers one of the parallels. In this way we can give each of the lengths s_m or d_m the appropriate direction at each point of the paper. For the determination

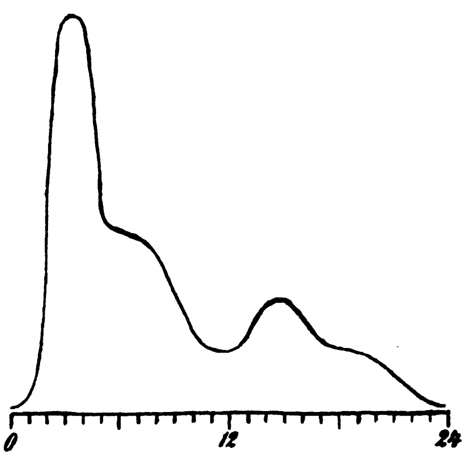


FIG. 114

of a coefficient with odd index, we start from the origin and draw the length d_1 in the prescribed direction. This makes an angle lx_1 with the positive x axis. We fix the endpoint of this by sticking a pin through the

paper, turn the strip through the angle lx_1, and then determine the endpoint of d_2. We again turn the strip about this point through an angle lx_1, etc. In this way we get the endpoint of the polygon, and consequently the coefficients a_l and b_l without any further drawing.

6. *Example*: Twenty four equidistant ordinates of the blood pressure curve (Fig. 114) are chosen, and the sums and differences of these are formed. The following values are obtained

s : 26, 60, 138, 136, 90, 68, 66, 60, 45, 29, 21, 19

d : −16, 0, 76, 82, 50, 32, 32, 30, 23, 17, 15, 15.

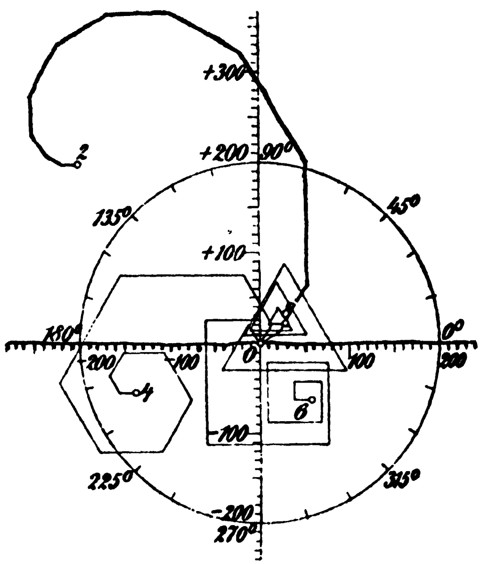


FIG. 115

The coefficients, up to the ninth, are constructed in Figs. 115 and 116. In Fig. 116, for the sums, the unit is 1/8 mm., while in Fig. 116, for the differences, we take the unit as 1/4 mm. From the figures we obtain the following values:

$a_1 = +\ 0.7;\ a_2 = -16.1;\ a_3 = -11.6;\ a_4 = -11.5;$

$b_1 = +23.8;\ b_2 = +16.5;\ b_3 = -\ 0.4;\ b_4 = -\ 4.6;$

$a_5 = -\ 0.2;\ a_6 = +\ 4.8;\ a_7 = +\ 4.3;\ a_8 = +\ 2.4;\ a_9 = +\ 0.4;$

$b_5 = -\ 9.0;\ b_6 = -\ 5.3;\ b_7 = +\ 0.3;\ b_8 = +\ 1.8;\ b_9 = +\ 2.2.$

7. The calculation of the sums (10) becomes completely mechanical if we use stencils, as suggested by *Hermann.*[3] Zipperer,[4] for example, has prepared such stencils for the case of 24 ordinates. The number 24 is

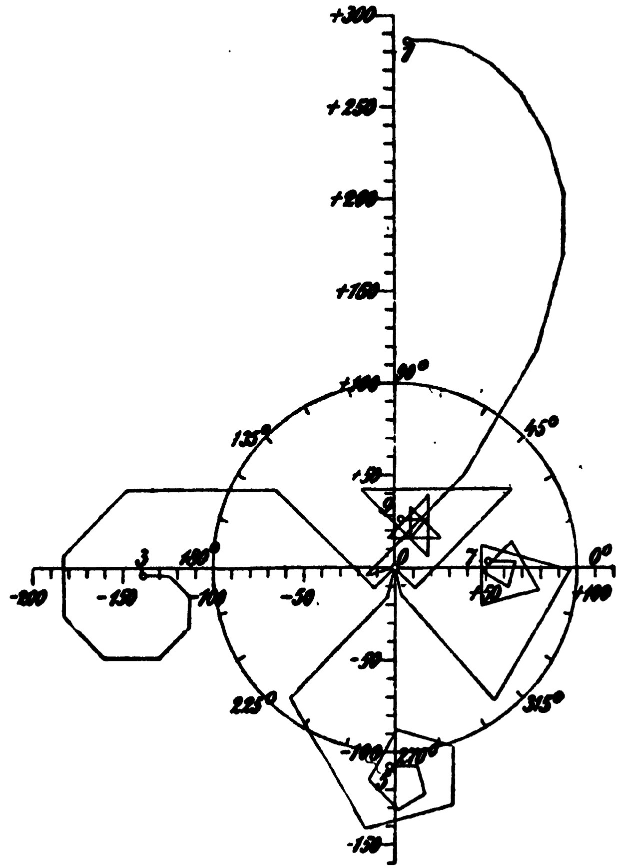


FIG. 116

useful because a large series of the cosine and sine terms appearing in the summand then becomes zero. The tables consist of a base table printed on tracing paper. The 24 ordinates and their products with sin ($\pm 75°$) = ± 0.966, sin ($\pm 60°$) = ± 0.866, sin ($\pm 45°$) = ± 0.707, sin ($\pm 30°$) = ± 0.5 and sin ($\pm 15°$) = ± 0.259 are plotted in the corresponding squares. The values which have been multiplied by the positive factors are on the right and those which have been multiplied by the negative factors are on the left. The squares with values not being considered are blackened initially. To avoid difficulties with the sign of the ordinate, we draw the x axis so low that all ordinates are positive; then this only changes the

value of a_0. For each of the coefficients a_1, b_1, $\cdots$, a_{11}, b_{11}, a special sheet is prepared with colored regions for the products which occur in the sum for the calculation of the coefficients involved, on the left or right side according as the trigonometric function which appears as a factor is positive or negative. These tables are placed under the base table and the numbers lying above a colored field on the right side are added together. The same is done on the left side, and the difference of the two sums is formed. If we divided this difference by 12, we get the coefficient in question. The results for the above example are in good agreement with the values found graphically:

$$\begin{aligned}
a_1 &= +\ 0.706;\ a_2 = -16.535;\ a_3 = -11.590;\ a_4 = -11.416;\\
b_1 &= +23.871;\ b_2 = -16.680;\ b_3 = -\ 0.216;\ b_5 = -\ 4.616;\\
a_5 &= -\ 0.199;\ a_6 = +\ 4.833;\ a_7 = +\ 4.487;\ a_8 = +\ 2.178;\\
b_5 &= -\ 8.843;\ b_6 = -\ 5.333;\ b_7 = -\ 0.287;\ b_8 = +\ 1.875;\\
a_9 &= +\ 0.426;\ a_{10} = -\ 0.676;\ a_{11} = -\ 1.326;\\
b_9 &= +\ 2.193;\ b_{10} = +\ 1.237;\ b_{11} = +\ 0.955.
\end{aligned}$$

In this case we have $a_0 = +31.583$; $a_{12} = -0.583$.

8. Computation work is reduced if, *from the outset, we group the ordinates* which are multiplied in the sums of the equations (10) with the equal positive or negative values of the trigonometric function, and if we calculate the coefficients with the sums or differences thus obtained according to a scheme of calculation as was given by *Runge*[5] or *Whittaker*[6] for 12 and for 24 ordinates. Here $x_m = 2\pi m/4h$, therefore

$$\cos lx_{4h-m} = \cos l(4h - m)\frac{2\pi}{4h} = \cos(l\cdot 2\pi - lx_m) = \cos lx_m\,,$$

$$\sin lx_{4h-m} = \sin l(4h - m)\frac{2\pi}{4h} = \sin(l\cdot 2\pi - lx_m) = -\sin lx_m\,.$$

The last two equations of (10) can also be written

$$\sum_{m=1}^{m=2h-1}(y_m + y_{4h-m})\cos lx_m + y_{2h}\cos l\pi + y_{4h}\cos 2\pi l = a_l\cdot 2h,$$

$$\sum_{m=1}^{m=2h-1}(y_m - y_{4h-m})\sin lx_m = b_l\cdot 2h,$$

since $\sin l\pi = \sin 2l\pi = 0$. If we now collect the coordinates according to the following scheme for sums and for differences (where we subtract the lower series from the upper),

$$\begin{array}{ccccccc} & y_1 & y_2 & y_3 & \cdots & y_{2h-1} & y_{2h} \\ y_{4h} & y_{4h-1} & y_{4h-2} & y_{4h-3} & \cdots & y_{2h+1} & \\ \hline \sigma_0 & \sigma_1 & \sigma_2 & \sigma_3 & \cdots & \sigma_{2h-1} & \sigma_{2h} \\ & \delta_1 & \delta_2 & \delta_3 & \cdots & \delta_{2h-1} & \end{array}$$

then the last two groups of equations (10) can be written

$$(11) \qquad \sum_{m=0}^{m=2h} \sigma_m \cos lx_m = 2ha_l \; ; \qquad \sum_{m=1}^{m=2h-1} \delta_m \sin lx_m = 2hb_l \; .$$

If we now observe further that

$$(11a) \qquad \begin{aligned} \cos lx_{2h-m} &= \cos l(2h - m)\frac{2\pi}{4h} = \cos(l\pi - lx_m) \\ &= \cos l\pi \cos lx_m - \sin l\pi \sin lx_m = \mp\cos lx_m \; , \\ \sin lx_{2h-m} &= \sin l(2h - m)\frac{2\pi}{4h} = \sin(l\pi - lx_m) \\ &= \sin l\pi \cos lx_m - \cos l\pi \sin lx_m = \pm\sin lx_m \; , \end{aligned}$$

where the upper sign holds for odd values of l, the lower for even values, then the equations (11) are

$$(11b) \qquad \begin{aligned} \sum_{m=0}^{m=h-1} (\sigma_m \mp \sigma_{2h-m}) \cos lx_m + \sigma_h \cos lx_h &= a_l \cdot 2h, \\ \sum_{m=1}^{m=h-1} (\delta_m \pm \delta_{2h-m}) \sin lx_m + \delta_h \sin lx_h &= b_l \cdot 2h. \end{aligned}$$

If we now combine the sums and differences again by the following scheme

$$(11c) \qquad \begin{array}{cccccc} \sigma_0 & \sigma_1 & \sigma_2 & \cdots & \sigma_{h-1} & \sigma_h \\ \sigma_{2h} & \sigma_{2h-1} & \sigma_{2h-2} & \cdots & \sigma_{h+1} & \\ \hline \sigma\sigma_0 & \sigma\sigma_1 & \sigma\sigma_2 & \cdots & \sigma\sigma_{h-1} & \sigma\sigma_h \\ \delta\sigma_0 & \delta\sigma_1 & \delta\sigma_2 & \cdots & \delta\sigma_{h-1} & \end{array} \qquad \begin{array}{ccccc} \delta_1 & \delta_2 & \cdots & \delta_{h-1} & \delta_h \\ \delta_{2h-1} & \delta_{2h-2} & \cdots & \delta_{h+1} & \\ \hline \sigma\delta_1 & \sigma\delta_2 & \cdots & \sigma\delta_{h-1} & \sigma\delta_h \\ \delta\delta_1 & \delta\delta_2 & \cdots & \delta\delta_{h-1} & \end{array} \; ,$$

then the equations (11) and the first two equations (10) may be written as follows, if we set $\sigma\delta_0 = \delta\delta_0 = \delta\delta_h = \delta\sigma_h = 0$;

a) for even values of l

(12a)
$$\left.\begin{aligned}\sum_{m=0}^{m=h} \sigma\sigma_m \cos lx_m &= a_l \cdot 2h \\ \sum_{m=0}^{m=h} \delta\delta_m \sin lx_m &= b_l \cdot 2h\end{aligned}\right\} (l = 2, 4, \cdots, 2h-2),$$

$$\sum_{m=0}^{m=h} \sigma\sigma_m = a_0 \cdot 4h, \qquad \sum_{m=0}^{m=h} \sigma\sigma_m(-1)^m = a_{2h} \cdot 4h;$$

b) for odd values of l

(12b)
$$\left.\begin{aligned}\sum_{m=0}^{m=h} \delta\sigma_m \cos lx_m &= a_l \cdot 2h \\ \sum_{m=0}^{m=h} \sigma\delta_m \sin lx_m &= b_l \cdot 2h\end{aligned}\right\} (l = 1, 3, \cdots, 2h-1).$$

In the calculation of this expression it is advisable always to calculate the coefficients b_l and b_{2h-l} , and also a_l and a_{2h-l} simultaneously, since in this way equal products always appear, only with different signs. If we write the products with one sign in one column and those with the other in a second, and add the columns, then we need only add or subtract these sums in order to get the two sets of coefficients.

9. The surest check for the correctness of the calculation is that we again calculate the function values used for the original computation, this time using the calculated coefficients. By addition and subtraction, it follows that

$$y_m = a_0 + a_1 \cos x_m + \cdots + a_{2h} \cos 2hx_m + b_1 \sin x_m + \cdots + b_{2h-1} \sin(2h-1)x_m$$

$$y_{4h-m} = a_0 + a_1 \cos x_m + \cdots + a_{2h} \cos 2hx_m - b_1 \sin x_m - \cdots - b_{2h-1} \sin(2h-1)x_m$$

(12c)
$$\frac{1}{2}\sigma_m = a_0 + a_1 \cos x_m + a_2 \cos 2x_m + \cdots + a_{2h} \cos 2hx_m$$
$$(m = 1, 2 \cdots 2h-1).$$
$$\frac{1}{2}\delta_m = b_1 \sin x_m + b_2 \sin 2x_m + \cdots + b_{2h-1} \sin(2h-1)x_m$$

Moreover,

(12d)
$$\sigma_m = a_0 + a_1 \cos x_m + a_2 \cos 2x_m + \cdots + a_{2h} \cos 2hx_m$$
$$(m = 0, 2h).$$

These equations are constructed exactly as were the equations (11) and the first two equations (10), if we introduce the sums of the ordinates σ_m. The quantities $1/2\sigma_m$ and $1/2\delta_m$ can then be computed from the coefficients according to the same scheme, as the $2h$-fold coefficients from the sums and differences of the coordinates.

To get a more rapid check, we can start from the minimum condition.

If the number of the coefficients is equal to the number of the ordinates, then by (9),

$$\begin{aligned}M = \sum_{m=1}^{m=4h} y_m^2 - 2h(2a_0^2 + a_1^2 + a_2^2 + \cdots \\ + a_{2h-1}^2 + 2a_{2h}^2 + b_1^2 + b_2^2 + \cdots + b_{2h-1}^2) = 0.\end{aligned} \tag{12e}$$

If the number of coefficients is smaller, then on the right side, we set a_m and b_m (which are not calculated) equal to zero, and get a measure for the accuracy achieved with the value of M. This equation can now be divided into two equations. Since

$$\begin{aligned}y_m^2 + y_{4h-m}^2 = \frac{1}{2}(y_m + y_{4h-m})^2 + \frac{1}{2}(y_m - y_{4h-m})^2 = \frac{1}{2}\sigma_m^2 + \frac{1}{2}\delta_m^2 \\ (m = 1, 2, \cdots, 2h-1),\end{aligned} \tag{12f}$$

the above equation can also be written

$$\begin{aligned}\sigma_0^2 + \frac{1}{2}\sum_{m=1}^{m=2h-1} \sigma_m^2 + \sigma_{2h}^2 + \frac{1}{2}\sum_{m=1}^{m=2h-1} \delta_m^2 - 4h(a_0^2 + a_{2h}^2) \\ - 2h\sum_{l=1}^{l=2h-1} a_l^2 - 2h\sum_{l=1}^{l=2h-1} b_l^2 = 0.\end{aligned} \tag{12g}$$

Now the coefficients a depend only on the quantities σ, while the coefficients b depend only on the δ. Consequently this equation can be broken up into two equations

$$\begin{aligned}\sigma_0^2 + \sigma_{2h}^2 + \frac{1}{2}\sum_{m=1}^{m=2h-1} \sigma_m^2 = 4h(a_0^2 + a_{2h}^2) + 2h\sum_{l=1}^{l=2h-1} a_l^2; \\ \frac{1}{2}\sum_{m=1}^{m=2h-1} \delta_m^2 = 2h\sum_{l=1}^{l=2h-1} b_l^2.\end{aligned} \tag{13}$$

If we want to use these formulas as a check on the accuracy of the calculation, we must calculate all the squares with the same absolute accuracy, since otherwise the errors of the larger squares outweigh those of the smaller ones.

10. In many cases, we are able to express the path of the function with sufficient accuracy with 12 *equidistant ordinates*. If we set $q = 12$, so that $h = 3$, then we obtain the following 12 equations from the equations (12) for the calculation of the coefficients. Here we replace the trigonometric functions by the equivalent functions in the first quadrant:

$$
(14)\quad
\begin{aligned}
12a_0 &= \sigma\sigma_0 + \sigma\sigma_1 + \sigma\sigma_2 + \sigma\sigma_3 \\
6a_1 &= \delta\sigma_0 + \delta\sigma_1 \cos 30^0 + \delta\sigma_2 \cos 60^0 \\
6a_2 &= \sigma\sigma_0 - \sigma\sigma_3 + (\sigma\sigma_1 - \sigma\sigma_2) \cos 60^0 \\
6a_3 &= \delta\sigma_0 - \delta\sigma_2 \\
6a_4 &= \sigma\sigma_0 + \sigma\sigma_3 + (-\sigma\sigma_1 - \sigma\sigma_2) \cos 60^0 \\
6a_5 &= \delta\sigma_0 - \delta\sigma_1 \cos 30^0 + \delta\sigma_2 \cos 60^0 \\
12a_6 &= \sigma\sigma_0 - \sigma\sigma_1 + \sigma\sigma_2 - \sigma\sigma_3 \\
6b_1 &= \sigma\delta_1 \sin 30^0 + \sigma\delta_2 \sin 60^0 + \sigma\delta_3 \\
6b_2 &= (\delta\delta_1 + \delta\delta_2) \sin 60^0 \\
6b_3 &= (\sigma\delta_1 - \sigma\delta_3) \\
6b_4 &= (\delta\delta_1 - \delta\delta_2) \sin 60^0 \\
6b_5 &= \sigma\delta_1 \sin 30^0 - \sigma\delta_2 \sin 60^0 + \sigma\delta_3 \, .
\end{aligned}
$$

These calculations can be carried out by means of the scheme below,[7] which is self explanatory, except that it should be observed that in the places characterized by $\sigma\sigma_m \delta\sigma_m$, etc., we do not have these values

Calculation Scheme

	Ordinates	Sums	Differences
	$y_1 \; y_2 \; y_3 \; y_4 \; y_5 \; y_6$	$\sigma_0 \; \sigma_1 \; \sigma_2 \; \sigma_3$	$\delta_1 \; \delta_2 \; \delta_3$
	$y_{12} \; y_{11} \; y_{10} \; y_9 \; y_8 \; y_7$	$\sigma_6 \; \sigma_5 \; \sigma_4$	$\delta_5 \; \delta_4$
Sums	$\sigma_0 \; \sigma_1 \; \sigma_2 \; \sigma_3 \; \sigma_4 \; \sigma_5 \; \sigma_6$	$\sigma\sigma_0 \; \sigma\sigma_1 \; \sigma\sigma_2 \; \sigma\sigma_3$	$\sigma\delta_1 \; \sigma\delta_2 \; \sigma\delta_3$
Differences	$\delta_1 \; \delta_2 \; \delta_3 \; \delta_4 \; \delta_5$	$\delta\sigma_0 \; \delta\sigma_1 \; \delta\sigma_2$	$\delta\delta_1 \; \delta\delta_2$

	Cosine terms			
$\cos 0^\circ = 1 \cdots\cdots$	$\sigma\sigma_0\sigma\sigma_1$ $\sigma\sigma_2\sigma\sigma_3$	$\delta\sigma_0$	$+\sigma\sigma_0-\sigma\sigma_3$	$\delta\sigma_0-\delta\sigma_2$
$\cos 30^\circ = 1 - 0.134$		$\delta\sigma_1$		
$\cos 60^\circ = 0.5 \cdots$		$\delta\sigma_2$	$-\sigma\sigma_2+\sigma\sigma_1$	
Sums $\cdots\cdots\cdots$	I II	I II	I II	I II
Sums I + II $\cdots$	$12a_0$	$6a_1$	$6a_2$	$6a_3$
Difference I − II $\cdot$	$12a_6$	$6a_5$	$6a_4$	

$2\sigma_0^2$	
σ_1^2	δ_1^2
σ_2^2	δ_2^2
σ_3^2	δ_3^2
σ_4^2	δ_4^2
σ_5^2	δ_5^2
$2\sigma_6^2$	
$[\sigma^2]$	$[\delta^2]$
$\left[\frac{1}{3}(6a)^2\right]$	$\frac{1}{3}[(6b^2)]$

	Sine terms		
$\sin 30 = 0.5 \cdots$	$\sigma\delta_1$		
$\sin 60^\circ = 1 - 0.134$	$\sigma\delta_2$	$\delta\delta_1\delta\delta_2$	
$\sin 90^\circ = 1 \cdots\cdots$	$\sigma\delta_3$		$\sigma\delta_1-\sigma\delta_3$
Sums $\cdots\cdots\cdots$	I II	I II	I II
Sum I + II $\cdots\cdots$	$6b_1$	$6b_2$	$6b_3$
Difference I − II $\cdot$	$6b_5$	$6b_4$	

$$
\begin{aligned}
y = a_0 &+ a_1\cos x + a_2\cos 2x \\
&+ a_3\cos 3x + a_4\cos 4x \\
&+ a_5\cos 5x + a_6\cos 6x \\
&+ b_1\sin x \ + b_2\sin 2x \\
&+ b_3\sin 3x \ + b_4\sin 4x \\
&+ b_5\sin 5x.
\end{aligned}
$$

	$(6a)^2$	$(6b^2)$	$(6a)^2+(6b)^2$	$(6r)^2:36$	r	$\operatorname{tg}\varphi = \frac{a}{b}$	Quadrant	$\bar{\varphi}$	
0	$\frac{1}{2}(12a_0)^2$								
1	$(6a_1)^2$	$(6b_1)^2$	$(6r_1)^2$	r_1^2	r_1	$\operatorname{tg}\varphi_1$		$\bar{\varphi}_1$	φ_1
2	$(6a_2)^2$	$(6b_2)^2$	$(6r_2)^2$	r_2^2	r_2	$\operatorname{tg}\varphi_2$		$\bar{\varphi}_2$	φ_2
3	$(6a_3)^2$	$(6b_3)^2$	$(6r_3)^2$	r_3^2	r_3	$\operatorname{tg}\varphi_3$		$\bar{\varphi}_3$	φ_3
4	$(6a_4)^2$	$(6b_4)^2$	$(6r_4)^2$	r_4^2	r_4	$\operatorname{tg}\varphi_4$		$\bar{\varphi}_4$	φ_4
5	$(6a_5)^2$	$(6b_5)^2$	$(6r_5)^2$	r_5^2	r_5	$\operatorname{tg}\varphi_5$		$\bar{\varphi}_5$	φ_5
6	$\frac{1}{2}(12a_6)^2$								
	$[(6a)^2]$	$[(6b)^2]$							

	Ordinates
	$-2.1-4.1-7.1-7.6-4.6-0.4$
	$0+3.1+7.3+7.2+4.2+2.2$
Sums	$0+1+3.2+0.1-3.4-2.4-0.4$
Differences	$-5.2-11.4-14.3-11.8-6.8$

	Sums	Differences
	$-0+1+3.2+0.1$	$-5.2-11.4-14.3$
	$-0.4-2.4-3.4$	$-6.8-11.8$
Sums	$-0.4-1.4-0.2+0.1$	$-12-23.2-14.3$
Differences	$+0.4+3.4+6.6$	$+1.6+0.4$

Cosine terms								
$\cos 0^0 = 1$	-0.4	-1.4						
	-0.2	$+0.1$	$+0.4$		-0.4	-0.1	$+0.4$	-6.6
$\cos 30^0 = 1 - 0.134$				$+2.94$				
$\cos 60^0 = 0.5$ . .			$+3.3$		$+0.1$	-0.7		
Sums	-0.6	-1.3	$+3.7$	$+2.94$	-0.3	-0.8	$+0.4$	-6.6
Sum I + II . . .	$12a_0 = -1.9$		$6a_1 = +6.64$		$6a_2 = -1.1$		$6a_3 = -6.2$	
Difference I − II	$12a_6 = +0.7$		$6a_5 = +0.76$		$6a_4 = +0.5$			

	Sine terms					
$\sin 30^0 = 0.5$	-6					
$\sin 60^0 = 1 - 0.134$.		-20.09	$+1.39$	$+0.35$		
$\sin 90^0 = 1$	-14.3				-12	$+14.3$
Sums	-20.3	-20.09	$+1.39$	$+0.35$	-12	$+14.3$
Sum I + II	$6b_1 = -40.39$		$6b_2 = +1.74$		$6b_3 = +2.3$	
Difference I − II . .	$6b_5 = -0.21$		$6b_4 = +1.04$			

$$
\begin{aligned}
y = -0.158 &+ 1.107 \cos \varphi - 6.732 \sin \varphi \\
&- 0.183 \cos 2\varphi + 0.290 \sin 2\varphi \\
&- 1.033 \cos 3\varphi + 0.383 \sin 3\varphi \\
&+ 0.083 \cos 4\varphi + 0.173 \sin 4\varphi \\
&+ 0.127 \cos 5\varphi - 0.035 \sin 5\varphi \\
&+ 0.056 \cos 6\varphi
\end{aligned}
$$

0	
1.00	27.04
10.24	129.96
0.01	204.49
11.56	139.16
5.76	46.24
0.32	
28.89	546.89
28.88	546.93

	$(6a)^2$	$(6b)^2$	$(6a)^2 + (6b)^2$	$(6r)^2{:}36$	r	$\operatorname{tg}\varphi = \frac{a}{b}$	Quadrant	$\bar{\varphi}$	φ
0	1.81								
1	44.09	1631.35	1675.44	46.540	6.822	−0.164	2	9.3	170.7
2	1.21	3.03	4.24	0.118	0.344	−0.632	4	32.3	327.7
3	38.44	5.29	43.73	1.215	1.102	−2.693	4	69.6	290.4
4	0.25	1.08	1.33	0.037	0.192	+0.481	1	25.6	25.6
5	0.58	0.04	0.62	0.017	0.130	−3.620	2	74.5	105.5
6	0.25								
	86.63	1640.79							

themselves, but their products with the trigonometric function values appearing at the beginning of the row. All calculations except the multiplication with sin 60° and cos 30° can be carried out mentally. We use the slide rule with these latter multiplications. However, we do not multiply by 0.866 but, because of the greater accuracy to be obtained, we multiply by 0.134 and subtract this value from the initial value. The calculation of the coefficients can be made with a single setting of the slide rule. In addition the scheme still has space for the check given under (13). Finally, a scheme is also given for the calculation of the cosine and sine series in a series of the form $\sum r_l \sin (l\varphi + \varphi_l)$ which therefore has the amplitude $r_l = (a_l^2 + b_l^2)^{\frac{1}{2}}$ and the phase angle $\varphi_l = \operatorname{arc\,tg} a_l/b_l$ for the individual terms.

11. *Example*: From the oscillogram of an alternating current wave, the following 12 ordinates are taken, measured in millimeters:

$$
\begin{array}{lllll}
 & y_1 = -2.1 & y_4 = -7.6 & y_7 = +2.2 & y_{10} = +7.3 \\
\text{(14a)} & y_2 = -4.1 & y_5 = -4.6 & y_8 = +4.2 & y_{11} = +3.1 \\
 & y_3 = -7.1 & y_6 = -0.4 & y_9 = +7.2 & y_{12} = 0.
\end{array}
$$

(See page 347.) By use of the amplitude and the phase angle, we get

$$
\begin{aligned}
y = -0.158 &+ 6.822 \sin (\varphi + 170.7) + 0.344 \sin (2\varphi + 327.7) \\
\text{(14b)} \qquad &+ 1.102 \sin (3\varphi + 290.4) + 0.192 \sin (4\varphi + 25.6) \\
&+ 0.130 \sin (5\varphi + 105.5) + 0.056 \cos 6\varphi.
\end{aligned}
$$

12. In many cases it is sufficient to determine the *first coefficients of the expansion.* If we pick out ordinates at arbitrary places on the graph of the function, we can determine the coefficients in

$$
\begin{aligned}
\bar{y} = a_0 &+ a_1 \cos x + a_2 \cos 2x + a_3 \cos 3x \\
&+ b_1 \sin x + b_2 \sin 2x + b_3 \sin 3x.
\end{aligned}
$$

This method, given by Thompson,[8] requires only addition, and no multiplication. From the equations (14) we get, for $q = 12$,

$$
\begin{aligned}
a_0 &= \frac{1}{12} \sum_{m=1}^{m=12} y_m , \\
\text{(15)} \qquad a_3 &= \frac{1}{6} (-y_2 + y_4 - y_6 + y_8 - y_{10} + y_{12}), \\
b_3 &= \frac{1}{6} (y_1 - y_3 + y_5 - y_7 + y_9 - y_{11}).
\end{aligned}
$$

If we now substitute successively in $\bar{y}$ the values $\pi/2$, $3\pi/2$, 0 and π, we get

$$
\begin{aligned}
y_3 &= a_0 - a_2 + b_1 - b_3 , \\
y_9 &= a_0 - a_2 - b_1 + b_3 , \\
\text{(16)} \qquad y_0 &= a_0 + a_1 + a_2 + a_3 , \\
y_6 &= a_0 - a_1 + a_2 - a_3 .
\end{aligned}
$$

By subtraction of the first two equations, we get

$$(17) \qquad b_1 = \frac{1}{2}(y_3 - y_9) + b_3 .$$

By subtraction of the last two,

$$(18) \qquad a_1 = \frac{1}{2}(y_0 - y_6) - a_3 ,$$

and by addition of the last two equations and subtraction of the first two, we find

$$(19) \qquad a_2 = \frac{1}{4}(y_0 - y_3 + y_6 - y_9).$$

To calculate b_2, we must start out from another interval division. If we denote the function values for the arguments $\pi/4$, $3\pi/4$, $5\pi/4$, $7\pi/4$ by $y_{3/2}$, $y_{9/2}$, $y_{15/2}$, $y_{21/2}$, and substitute for $h = 2$ in the equation (10) determining b_2, we find

$$(20) \qquad b_2 = \frac{1}{4}(y_{3/2} - y_{9/2} + y_{15/2} - y_{21/2}).$$

By this formula, we get in the above example,

$$(20a) \qquad \begin{aligned} &a_0 = -0.158; \quad a_3 = -1.033; \quad b_3 = +0.383; \quad b_1 = -6.767; \\ &a_1 = +1.233; \quad a_2 = -0.125; \quad b_2 = +0.175, \end{aligned}$$

in satisfactory agreement with the values found in the previous section.

13. For more accurate analysis it is advisable to use 24 ordinates. For this number of ordinates there exist printed *calculation formulas of Runge and Emde.*[9]. In this case the equations (12) become, if we transform the angle functions directly,

$$\begin{aligned} 24a_0 &= \sigma\sigma_0 + \sigma\sigma_2 + \sigma\sigma_4 + \sigma\sigma_6 \\ &\quad + \sigma\sigma_1 + \sigma\sigma_3 + \sigma\sigma_5 \\ 12a_1 &= \delta\sigma_0 + \delta\sigma_2 \cos 30^\circ + \delta\sigma_4 \cos 60^\circ \\ &\quad + \delta\sigma_1 \cos 15^\circ + \delta\sigma_3 \cos 45^\circ + \delta\sigma_5 \cos 75^\circ \\ 12a_2 &= \sigma\sigma_0 - \sigma\sigma_6 + (\sigma\sigma_2 - \sigma\sigma_4) \cos 60^\circ \\ &\quad + (\sigma\sigma_1 - \sigma\sigma_5) \cos 30^\circ \end{aligned}$$

$$12a_3 = \delta\sigma_0 - \delta\sigma_4$$
$$+ (\delta\sigma_1 - \delta\sigma_3 - \delta\sigma_5)\cos 45^0$$

$$12a_4 = \sigma\sigma_0 + \sigma\sigma_6 + (-\sigma\sigma_2 - \sigma\sigma_4)\cos 60^0$$
$$- \sigma\sigma_3 + (\sigma\sigma_1 + \sigma\sigma_5)\cos 60^0$$

$$12a_5 = \delta\sigma_0 - \delta\sigma_2 \cos 30^0 + \delta\sigma_4 \cos 60^0$$
$$+ \delta\sigma_1 \cos 75^0 - \delta\sigma_3 \cos 45^0 + \delta\sigma_5 \cos 15^0$$

$$12a_6 = \sigma\sigma_0 - \sigma\sigma_2 + \sigma\sigma_4 - \sigma\sigma_6$$

$$12a_7 = \delta\sigma_0 - \delta\sigma_2 \cos 30^0 + \delta\sigma_4 \cos 60^0$$
$$- \delta\sigma_1 \cos 75^0 + \delta\sigma_3 \cos 45^0 - \delta\sigma_5 \cos 15^0$$

$$12a_8 = \sigma\sigma_0 + \sigma\sigma_6 + (-\sigma\sigma_2 - \sigma\sigma_4)\cos 60^0$$
$$+ \sigma\sigma_3 + (-\sigma\sigma_1 - \sigma\sigma_5)\cos 60^0$$

$$12a_9 = \delta\sigma_0 - \delta\sigma_4$$
$$+ (-\delta\sigma_1 + \delta\sigma_3 + \delta\sigma_5)\cos 45^0$$

$$12a_{10} = \sigma\sigma_0 - \sigma\sigma_6 + (\sigma\sigma_2 - \sigma\sigma_4)\cos 60^0$$
$$+ (-\sigma\sigma_1 + \sigma\sigma_5)\cos 30^0$$

$$12a_{11} = \delta\sigma_0 + \delta\sigma_2 \cos 30^0 + \delta\sigma_4 \cos 60^0$$
$$- \delta\sigma_1 \cos 15^0 - \delta\sigma_3 \cos 45^0 - \delta\sigma_5 \cos 75$$

(21)

$$24a_{12} = \sigma\sigma_0 + \sigma\sigma_2 + \sigma\sigma_4 + \sigma\sigma_4$$
$$- \sigma\sigma_1 - \sigma\sigma_3 - \sigma\sigma_5$$

$$12b_1 = \sigma\delta_2 \sin 30^0 + \sigma\delta_4 \sin 60^0 + \sigma\delta_6$$
$$+ \sigma\delta_1 \sin 15^0 + \sigma\delta_3 \sin 45^0 + \sigma\delta_5 \sin 75^0$$

$$
\begin{aligned}
12b_2 &= (\delta\delta_2 + \delta\delta_4)\sin 60^0 \\
&\qquad + \delta\delta_3 + (\delta\delta_1 + \delta\delta_5)\sin 30^0 \\
12b_3 &= \sigma\delta_2 - \sigma\delta_6 \\
&\qquad + (\sigma\delta_1 + \sigma\delta_3 - \sigma\delta_5)\sin 45^0 \\
12b_4 &= (\delta\delta_2 - \delta\delta_4)\sin 60^0 \\
&\qquad + (\delta\delta_1 - \delta\delta_5)\sin 60^0 \\
12b_5 &= \sigma\delta_2 \sin 30^0 - \sigma\delta_4 \sin 60^0 + \sigma\delta_6 \\
&\qquad + \sigma\delta_1 \sin 75^0 - \sigma\delta_3 \sin 45^0 + \sigma\delta_5 \sin 15^0 \\
12b_6 &= \\
&\qquad \delta\delta_1 - \delta\delta_3 + \delta\delta_5 \\
12b_7 &= -\sigma\delta_2 \sin 30^0 + \sigma\delta_4 \sin 60^0 - \sigma\delta_6 \\
&\qquad + \sigma\delta_1 \sin 75^0 - \sigma\delta_3 \sin 45^0 + \sigma\delta_5 \sin 15^0 \\
12b_8 &= (-\delta\delta_2 + \delta\delta_4)\sin 60^0 \\
&\qquad + (\delta\delta_1 - \delta\delta_5)\sin 60^0 \\
12b_9 &= -\sigma\delta_2 + \sigma\delta_6 \\
&\qquad + (\sigma\delta_1 + \sigma\delta_3 - \sigma\delta_5)\sin 45^0 \\
12b_{10} &= (-\delta\delta_2 - \delta\delta_4)\sin 60^0 \\
&\qquad + \delta\delta_3 + (\delta\delta_1 + \delta\delta_5)\sin 30^0 \\
12b_{11} &= -\sigma\delta_2 \sin 30^0 - \sigma\delta_4 \sin 60^0 - \sigma\delta_6 \\
&\qquad + \sigma\delta_1 \sin 15^0 + \sigma\delta_3 \sin 45^0 + \sigma\delta_5 \sin 75^0.
\end{aligned}
$$

In the first half, the summands are listed which contain the ordinates with even index, while in the second half are those with ordinates of odd index. Now the summands from the ordinates with even index in the coefficients $a_0, \cdots, a_6$ and $b_1, \cdots, b_5$ agree exactly with the values of

the equations (14), if only we replace the index l by $2l$. The same summands are repeated in reverse order with the coefficients a_7 to a_{12} with the same sign, and with the coefficients b_7 to b_{11} with the opposite sign. These portions of the sums can therefore be calculated by use of the scheme given in Sec. 9. From the ordinates $y_2, y_4, y_6, \cdots, y_{24}$ we then get values which may be denoted by $12\alpha_0, 6\alpha_1, \cdots, 6\alpha_5, 12\alpha_6, 6\beta_1, \cdots, 6\beta_5$.

Furthermore, it is then only a question of the calculation of the part of the sum which depends on the ordinates with odd index. These summands also repeat themselves—those from a_0 to a_5 with opposite signs in reverse order to a_7 to a_{12}, those from b_1 to b_5 in reverse sequence with the same sign as those from b_7 to b_{11}. For the calculation of this part of the coefficients we return to the summation form. From the odd ordinates we get

$$12\bar{\alpha}_0 = \sum_{m=1}^{m=12} y_{2m-1} \qquad 12\bar{\beta}_6 = \sum_{m=1}^{m=12} y_{2m-1} \sin 6x_{2m-1},$$

$$(22) \quad \left.\begin{aligned} 6\bar{\alpha}_l &= \sum_{m=1}^{m=12} y_{2m-1} \cos lx_{2m-1} \\ 6\bar{\beta}_l &= \sum_{m=1}^{m=12} y_{2m-1} \sin lx_{2m-1} \end{aligned}\right\} (l = 1, \cdots, 5).$$

These expressions can be transformed. If we set

$$(23) \qquad x_{2m-1} = x_{2m-4} + x_3,$$

where $x_3 = \pi/4$, and introduce, for short,

$$12\alpha_l' = \sum_{m=1}^{m=12} y_{2m-1} \cos lx_{2m-4} \qquad (l = 0; 6)$$

$$(24) \quad \left.\begin{aligned} 6\alpha_l' &= \sum_{m=1}^{m=12} y_{2m-1} \cos lx_{2m-4} \\ 6\beta_l' &= \sum_{m=1}^{m=12} y_{2m-1} \sin lx_{2m-4} \end{aligned}\right\} (l = 1; 2 \cdots 5),$$

sums which can be calculated from the scheme for 12 ordinates, we get

$$12\bar{\alpha}_0 = 12\alpha_0' \qquad 12\bar{\beta}_0 = 0$$

$$(25) \quad 6\bar{\alpha}_1 = 3(2)^{1/2}(\alpha_1' - \beta_1') \qquad 6\bar{\beta}_1 = 3(2)^{1/2}(\alpha_1' + \beta_1')$$

$$6\bar{\alpha}_2 = -6\beta_2' \qquad 6\bar{\beta}_2 = +6\alpha_2'$$

$$6\bar{\alpha}_3 = 3(2)^{1/2}(-\alpha_3' - \beta_3') \qquad 6\bar{\beta}_3 = 3(2)^{1/2}(\alpha_3' - \beta_3')$$

$$6\bar{\alpha}_4 = -6\alpha_4' \qquad 6\bar{\beta}_4 = -6\beta_4'$$

$$6\bar{\alpha}_5 = 3(2)^{1/2}(-\alpha_5' + \beta_5') \qquad 6\bar{\beta}_5 = 3(2)^{1/2}(-\alpha_5' - \beta_5')$$

$$12\bar{\alpha}_6 = 0 \qquad 12\bar{\beta}_6 = -12\alpha_6' .$$

To calculate the quantities α_l' and β_l' by means of the scheme for 12 ordinates, we must collect the ordinates in the following way:

$$\text{(25a)} \qquad \begin{array}{cccccccc} & y_5 & y_7 & y_9 & y_{11} & y_{13} & y_{15} & \\ y_3 & y_1 & y_{23} & y_{21} & y_{19} & y_{17} & & . \end{array}$$

Then we get

$$\sum_{m=1}^{m=12} y_{2m+3} \cos lx_{2m} = \sum_{m=1}^{m=12} y_{2m-1} \cos lx_{2m-4} = 12\alpha_l' \qquad (l = 0, 6)$$

$$\text{(26)} \qquad \left.\begin{aligned} \sum_{m=1}^{m=12} y_{2m+3} \cos lx_{2m} &= \sum_{m=1}^{m=12} y_{2m-1} \cos lx_{2m-4} = 6\alpha_l' \\ \sum_{m=1}^{m=12} y_{2m+3} \sin lx_{2m} &= \sum_{m=1}^{m=12} y_{2m-1} \sin lx_{2m-4} = 6\beta_l' \end{aligned}\right\} (l = 1 \cdots 5).$$

The change in index is allowed because of the periodicity of the trigonometric function and the function to be approximated.

Therefore, to get the coefficients for 24 ordinates, we apply the scheme in Sec. 10, first on the even, then on the odd ordinates. Furthermore, we then use the following arrangement:

$6\alpha_1'$ $6\beta_1'$	$6\alpha_3'$ $6\beta_3'$	$6\alpha_5'$ $6\beta_5'$
Σ_1 Δ_1	Σ_3 Δ_3	Σ_5 Δ_5
$6\bar{\alpha}_1 = \frac{1}{(2)^{1/2}} \Delta_1$	$6\bar{\alpha}_3 = -\frac{1}{(2)^{1/2}} \Sigma_3$	$6\bar{\alpha}_5 = -\frac{1}{(2)^{1/2}} \Delta_5$
$6\bar{\beta}_1 = \frac{1}{(2)^{1/2}} \Sigma_1$	$6\bar{\beta}_3 = +\frac{1}{(2)^{1/2}} \Delta_3$	$6\bar{\beta}_5 = -\frac{1}{(2)^{1/2}} \Sigma_5$

(26a)

	$12\alpha_0$	$6\alpha_1$	$6\alpha_2$	$6\alpha_3$	$6\alpha_4$	$6\alpha_5$	$6\alpha_6$
	$12\alpha_0'$	$6\bar{\alpha}_1$	$-6\beta_2'$	$6\bar{\alpha}_3$	$-6\alpha_4'$	$6\bar{\alpha}_5$	
Sum	$24a_0$	$12a_1$	$12a_2$	$12a_3$	$12a_4$	$12a_5$	$12a_6$
Difference	$24a_{12}$	$12a_{11}$	$12a_{6}$	$12a_9$	$12a_8$	$12a_7$	
	$6\bar{\beta}_1$	$6\alpha_2'$	$6\bar{\beta}_3$	$-6\beta_4'$	$6\bar{\beta}_5$	$-12\alpha_6'$	
	$6\beta_1$	$6\beta_2$	$6\beta_3$	$6\beta_4$	$6\beta_5$		
Sum	$12b_1$	$12b_2$	$12b_3$	$12b_4$	$12b_5$	$12b_6$	
Difference	$12b_{11}$	$12b_{10}$	$12b_9$	$12b_8$	$12b_7$.		

For the reduction of the series thus set up into a series of the form $\sum r_m \cos (l\varphi - \varphi_m)$, the corresponding scheme is valid, which was given on page 347 for 12 ordinates. Likewise, we can apply the calculation checks given in Sec. 9. An example may easily be worked through by the reader.

Finally, an apparatus should be mentioned which forms the products entering into the sums (7) mechanically, and also carries out the summation mechanically. This is the *analyzer of Michelson and Stratton*[10] which, for example, is carried out for 80 ordinates. Because of the relations given at the beginning of Sec. 8, this apparatus can be used not only for analysis, i.e., for the calculation of the Fourier coefficients, but also for the synthesis, i.e., the calculation of the function values from the coefficients.[11]

NOTES

1. v. Sanden, *Z. f. Math. u. Phys.* **61** (1912-13), p. 430.
2. Groeneveld, *Z. f. angew. Math. u. Mech.* 6 (1926), p. 253.
3. *Pflügers Archiv.* (1890), p. 45.
4. Zipperer, *Tafeln zur harmonischen Analyse periodischer Funktionen* (Berlin, 1922). (The table for A_4 contains numerous errors.) Earlier tables were compiled by Pollak.
5. Runge, *Z. f. Math. u. Phys.* 78 (1902).
6. Whittaker and Robinson, *The Calculus of Observations*, 2nd ed. (London, 1926), Ch. X.
7. Runge, *Z. f. Math. u. Phys.* 48 (1903), p. 443.
8. S. P. Thompson, *Proc. Phys. Soc. London* **23** (1911), p. 334.
9. Runge, *Erläuterung des Rechnungsformulars zur Zerlegung einer empirisch gegebenen periodischen Funktion in Sinuswellen* (Braunschweig, 1913).
10. *The American Journal of Science* (4) 5 (1898), No. 25, 1.
11. Willers, *Math. Instrumente* (Berlin, 1926), Art. 17.

30. Representation by Exponential Functions.

1. The approximation by exponential functions comes into considera-

tion *in three cases*: first, wherever decay processes are involved, such as radioactive processes, cooling, etc.; second, in decay processes with damped oscillations, as is found in charging processes; third, with phenomena which arise by the superposition of purely periodic processes whose periods do not have integral ratios. We find the latter in the brightness fluctuations of variable stars, the ebb and flow of tides, or in the oscillation processes in the production of vowels and half vowels. The distinction for the representation of such processes is only that in the first case we deal with exponential functions with real exponents, while in the second and third cases the exponents are complex or purely imaginary. If by γ_m we indicate arbitrary real or complex quantities, we can form the series

$$y(x) = C_1Ce^{\gamma_1 x} + C_2e^{\gamma_2 x} + C_3e^{\gamma_3 x} + \cdots .$$

This series is more flexible than those used previously, because the individual approximation functions themselves contain the constants γ_m still to be suitably chosen. First, if these constants are fixed, the functions to be used for the approximation are found, and then, if necessary, the coefficients C are determined by the principle of the minimum.

2. It is of great importance for the application not to carry along too few exponential functions. In many cases we shall so control the process to be represented that we can determine the number of the terms of the approximating function to be used. In other cases, we shall be able to determine the number from the path of the curve. In the most important applications, three or four terms will be sufficient. Therefore we shall *limit the discussion here to the development of* 3 *summands*. An extension to n terms follows directly.

Therefore we consider

$$\bar{y}(x) = C_1e^{\gamma_1 x} + C_2e^{\gamma_2 x} + C_3e^{\gamma_3 x}. \tag{1}$$

For simplicity, we assume that the observations are equidistant, and that the difference in the successive abscissas is h. The value of the mth ordinate of the function (1) is then

$$\bar{y}_m = C_1e^{\gamma_1[x_1+(m-1)h]} + C_2e^{\gamma_2[x_1+(m-1)h]} + C_3e^{\gamma_3[x_1+(m-1)h]}, \tag{1a}$$

or, if we use the abbreviations

$$e^{\gamma_1 h} = u_1 \qquad e^{\gamma_2 h} = u_2 \qquad e^{\gamma_3 h} = u_3$$

(1b)

$$C_1e^{\gamma_1[x_1+(m-1)h]} = f_1\,; \quad C_2e^{\gamma_2[x_1+(m-1)h]} = f_2\,; \quad C_3e^{\gamma_3[x_1+(m-1)h]} = f_3\,,$$

then for 4 equidistant measurements, we have

$$
(2)\qquad
\begin{aligned}
y_m &= f_1 + f_2 + f_3\\
y_{m+1} &= f_1u_1 + f_2u_2 + f_3u_3\\
y_{m+2} &= f_1u_1^2 + f_2u_2^2 + f_3u_3^2\\
y_{m+3} &= f_1u_1^3 + f_2u_2^3 + f_3u_3^3 .
\end{aligned}
$$

under the assumption that the actual value y agrees with the value $\bar{y}$ of the approximating function.

In many cases we do not know the zero line to which the process decays. Consequently there will be a term $u = 1$ in the above series, i.e., the corresponding $\gamma = 0$. We must therefore carry one additional term. Instead of the series (1), it is better to begin with

$$
(2a)\qquad \bar{y} = C_0 + C_1e^{\gamma_1 x} + C_2e^{\gamma_2 x} + C_3e^{\gamma_3 x}.
$$

To avoid the greater work of computation which arises in this case, we can form the difference of two successive ordinates for the elimination of C_0 . Then, instead of the equation (2) we get

$$
(3)\qquad
\begin{aligned}
\Delta_{m+(1/2)} &= f_1(u_1 - 1) + f_2(u_2 - 1) + f_3(u_3 - 1)\\
\Delta_{m+(3/2)} &= f_1(u_1 - 1)u_1 + f_2(u_2 - 1)u_2 + f_3(u_3 - 1)u_3\\
\Delta_{m+(5/2)} &= f_1(u_1 - 1)u_1^2 + f_2(u_2 - 1)u_2^2 + f_3(u_3 - 1)u_3^2\\
\Delta_{m+(7/2)} &= f_1(u_1 - 1)u_1^3 + f_2(u_2 - 1)u_2^3 + f_3(u_3 - 1)u_3^3 ,
\end{aligned}
$$

in which we must use five successive function values. The 4 equations (2) or (3) for the 3 quantities f_1 , f_2 , f_3 can only exist simultaneously if the determinant of the coefficients is zero. This gives a relation between the u_1 , u_2 , u_3 and the y or Δ, which we also get if we multiply the four equations with the symmetric functions of these three quantities, taken with corresponding signs, the first with $s_3 = -u_1u_2u_3$, the second with $s_2 = u_1u_2 + u_2u_3 + u_3u_1$ and the third with $s_1 = -(u_1 + u_2 + u_3)$ and then add all equations. We then get from (3) the equation

$$
(4)\qquad \Delta_{m+(1/2)}s_3 + \Delta_{m+(3/2)}s_2 + \Delta_{m+(5/2)}s_1 + \Delta_{m+(7/2)} = 0.
$$

If we start out with the equation (2), then the y appear in place of the Δ. The equation (4) is valid for each value of m; then n equidistant function values are present, and we get $n - 3$ equations if we start from (2), $n - 4$ if we start from (3):

$$
\begin{aligned}
&\Delta_{1/2}s_3 + \Delta_{3/2}s_2 + \Delta_{5/2}s_1 + \Delta_{7/2} = \\
&\Delta_{3/2}s_3 + \Delta_{5/2}s_2 + \Delta_{7/2}s_1 + \Delta_{9/2} = 0 \\
&\Delta_{5/2}s_3 + \Delta_{7/2}s_2 + \Delta_{9/2}s_1 + \Delta_{11/2} = 0 \\
(5)\qquad &\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot \\
&\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot \\
&\Delta_{n-(7/2)}s_3 + \Delta_{n-(5/2)}s_2 + \Delta_{n-(3/2)}s_1 + \Delta_{n-(1/2)} = 0.
\end{aligned}
$$

For the determination of the s in this case therefore, at least 7 function values must be considered. In general, more values will be available. Then we shall regard these equations, no one of which will be satisfied exactly (and consequently there will be a small value ϵ_m different from zero on the right side of the mth equation), as error equations, and we can calculate the *normal equations* from according to the method of least squares (Art. 25):

$$
\begin{aligned}
&[\Delta_{m+(1/2)}\Delta_{m+(1/2)}]s_3 + [\Delta_{m+(1/2)}\Delta_{m+(3/2)}]s_2 + [\Delta_{m+(1/2)}\Delta_{m+(5/2)}]s_1 \\
&\qquad\qquad + [\Delta_{m+(1/2)}\Delta_{m+(7/2)}] = 0 \\
&[\Delta_{m+(3/2)}\Delta_{m+(1/2)}]s_3 + [\Delta_{m+(3/2)}\Delta_{m+(3/2)}]s_2 + [\Delta_{m+(3/2)}\Delta_{m+(5/2)}]s_1 \\
&\qquad\qquad + [\Delta_{m+(3/2)}\Delta_{m+(7/2)}] = 0 \\
(6)\qquad &[\Delta_{m+(5/2)}\Delta_{m+(1/2)}]s_3 + [\Delta_{m+(5/2)}\Delta_{m+(3/2)}]s_2 + [\Delta_{m+(5/2)}\Delta_{m+(5/2)}]s_1 \\
&\qquad\qquad + [\Delta_{m+(5/2)}\Delta_{m+(7/2)}] = 0 \\
&[\Delta_{m+(7/2)}\Delta_{m+(1/2)}]s_3 + [\Delta_{m+(7/2)}\Delta_{m+(3/2)}]s_2 + [\Delta_{m+(7/2)}\Delta_{m+(5/2)}]s_1 \\
&\qquad\qquad + [\Delta_{m+(7/2)}\Delta_{m+(7/2)}] = [\epsilon\epsilon].
\end{aligned}
$$

In the sums, m runs from 1 to $n - 4$. From the last equation we get (by 25(4)) the *mean error* of an equation $m = (([\epsilon\epsilon])/(n - 7))^{\frac{1}{2}}$ and from this the mean error of an ordinate difference

$$(7)\qquad m_\Delta = \left(\frac{[\epsilon\epsilon]}{(n-7)(1 + s_1^2 + s_2^2 + s_3^2)}\right)^{1/2}.$$

If we started from the equations (2), we would have $n - 6$ equations, since we have one more equation, and the corresponding expression (7)

gives the error of a single measurement. If this value does not become small, we must repeat the same calculation with an approximating function with more terms.

3. From the normal equations we get the *symmetric functions* s_1, s_2, s_3, the coefficients of the equation of third degree,

$$x^3 + s_1x^2 + s_2x + s_3 = 0. \tag{8}$$

The roots of this equation are the desired values u_1, u_2, u_3. If this equation has only real roots, then the coefficients are determined from

$$\gamma_1 = \frac{1}{h} \ln u_1 ; \qquad \gamma_2 = \frac{1}{h} \ln u_2 ; \qquad \gamma_3 = \frac{1}{h} \ln u_3 . \tag{9}$$

On the other hand, if it has one real and two complex conjugate roots u_1, $u_{2,3} = v \pm iw$, then the two remaining γ are also complex conjugates $\gamma_{2,3} = \alpha + i\beta$, and we have

$$\alpha = \frac{1}{2h} \ln(v^2 + w^2), \qquad \beta = \frac{1}{h} \operatorname{arc\,tg} \frac{w}{v} \pm \frac{2\kappa\pi}{h}. \tag{10}$$

In this case it is better to use the approximating function in the form

$$\bar{y} = C_1e^{\gamma_1 x} + Ae^{\alpha x} \cos \beta x + Be^{\alpha x} \sin \beta x. \tag{11}$$

The functions appearing in the terms of the approximating function are determined by this, and frequently the problem is then completely solved, since we are often interested only in the *determination of the half life, the damping decrement, the period*, etc.

If in addition we want to represent the path of the function, we must determine the coefficients by the minimum condition. This will be omitted here.[1]

4. *Example*: In a body cooling to constant temperature, the temperature T was measured x minutes after the beginning of the cooling. The initial and final temperatures are sought.

x	2	4	6	8	10	12	14	16	18	20
T	92.4	86.2	80.5	75.2	70.3	65.8	61.6	57.7	54.1	50.8

This is of the form $T = A + Be^{\gamma x}$. From this we get 8 equations of the form $\Delta_n u + \Delta_{n-1} = 0$ from which the mean value $u = 0.924 \pm 0.002$ is obtained, i.e., $\gamma = 1/2 \ln 0.924$. If we take 10 as the base number for more convenient calculation, $\bar{\gamma} = 1/2 \log 0.924 = -0.0172$. In this way, the approximating function

$$T = A + B10^{-0.0172x}$$

is obtained. If we substitute the values of x and T in this equation, we get 7 error equations, from which the normal equations [25(11)] are then given:

$$10A + 6.645B - 694.6 = 0$$

$$6.645A + 4.6403B - 481.445 = 0$$

$$694.6A + 481.445B - 50006.4 = [\epsilon\epsilon].$$

From these, the values $A = 10.66$; $B = 88.487$; $[\epsilon\epsilon] = 0.1$ are calculated by the methods of 23.3. We therefore have the approximate representation

$$T = 10.66 + 88.49 \times 10^{-0.0172x}.$$

This gives 99°.15 as the initial temperature, and 10°.66 as the final temperature.

5. If we have to analyze *purely periodic processes* of unknown period, then the exponentials to be chosen are imaginary. In the case of the superposition of two oscillations of different period, we then have to form the series

$$\bar{y} = A_1e^{\alpha ix} + A_2e^{-\alpha ix} + B_1e^{\beta ix} + B_2e^{-\beta ix}.$$

If the observations are equidistant with interval h, we have $u_1 = e^{\alpha ih}$, $u_2 = e^{-\alpha ih}$, $u_3 = e^{\beta ih}$, $u_4 = e^{-\beta ih}$, i.e., $u_1 = 1/u_2$, $u_3 = 1/u_4$. The equation for the determination of u, corresponding to (8)

(8a) $$x^4 + s_1x^3 + s_2x^2 + s_3x + s_4 = 0$$

must therefore become a reciprocal equation, i.e., $s_4 = 1$, $s_1 = s_3$. The determination of the values of the roots can then be reduced to quadratic equations. Also the corresponding equation (4) for the determination of the symmetric functions becomes simpler, because of the relation $s_4 = 1$, $s_3 = s_2$,

(4a) $$\Delta_{m+(9/2)} + \Delta_{m+(1/2)} + (\Delta_{m+(7/2)} + \Delta_{m+(3/2)})s_1 + \Delta_{m+(5/2)}s_2 = 0.$$

The equations are treated in the same way if we have several periods to combine.

6. *Example*: Whittaker and Robinson[2] analyze by systematic tests the period of a variable star. For this purpose they use 600 function values which give the magnitude of the star at midnight on 600 successive days. The unit is arbitrarily chosen. These observations

are represented in Fig. 117. From the path of the curve we see that it is evidently a question of the superposition of two periodic processes.[3] We therefore form the series (4a). We could form 595 such equations. To shorten the work of calculation and still use as many

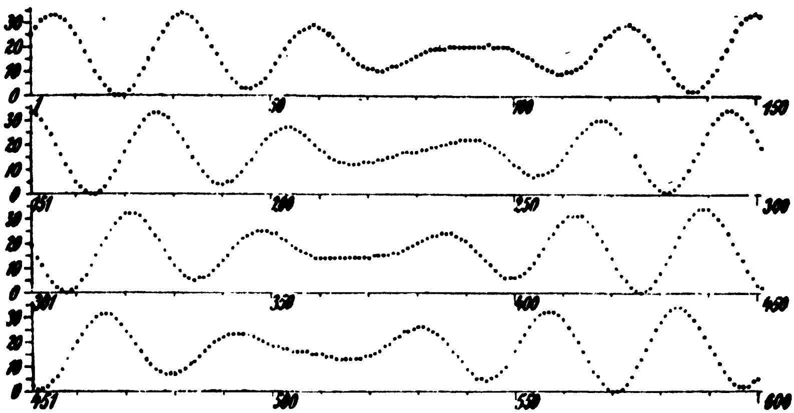


FIG. 117

of the data as possible, 78 equations are formed, in which $h = 10$. We start out from the 1st, 8th, 15th, $\cdots$ function values. The first equations would then be

$$(13 - 10 + 22 - 25) + s_1(10 - 33 + 2 - 22) + s_2(33 - 2)$$

$$= 31s_2 - 43s_1 + 0 = 0$$

$$(28 - 5 + 0 - 31) + s_1(5 - 20 + 26 - 0) + s_2(20 - 26)$$

$$= -6s_2 + 11s_1 - 8 = 0$$

. . . .

These equations would be taken as error equations and the normal equations formed from these by the method of least squares [25(11)]. These become

$$19657s_2 - 26548s_1 - 1830 = 0$$

$$26548s_2 - 37616s_1 + 2541 = 0$$

$$-1830s_2 - 2541s_1 + 14504 = [\epsilon\epsilon].$$

From these we get

$$s_2 = 3.9367496, \qquad s_1 = 2.8459652, \qquad [\epsilon\epsilon] = 68.151.$$

As mean error of an equation we find from these

$$m = \left(\frac{68.15}{76}\right)^{1/2} = 0.946,$$

and as the mean error of a difference,

$$m_\Delta = \left(\frac{68.15}{76 \times 24.5}\right)^{1/2} = 0.191.$$

The reciprocal equation for the determination of u therefore becomes

$$u^4 + 2.8459652u^3 + 3.9367496u^2 + 2.8459652u + 1 = 0.$$

If we introduce $y = u + 1/u$ here, we get a quadratic equation for y

$$y^2 + 2.8459652y + 1.9367496 = 0,$$

which has the two roots $y_1 = -1.7198488$, $y_2 = -1.1261162$. The values of u are then calculated from the two quadratic equations

$$u^2 + 1.1261162u + 1 = 0, \qquad u^2 + 1.7198488u + 1 = 0$$

and are

$$u_{1,2} = -0.563058 \pm 0.8264173i; \, u_{3,4} = -0.8599244 \pm 0.5104214i.$$

From these we obtain angles in the second or third quadrants, or such angles which are larger by $k \times 360°$; however, we see immediately from the figure that these angles are unimportant. Here we are only concerned with the angles in the second quadrant, namely

$$10\alpha = 124°.268 \qquad 10\beta = 149°.308$$

$$\alpha = 12°.427 \qquad \beta = 14°.931,$$

and from these we get the length of the period as 28.97 or 24.11 days.

To determine which of the various angles is important, 17 equations are set up for $h = 7$, starting out from the 1st, 35th, 70th, $\cdots$ function values. From these 17 error equations, the normal equations

$$2858s_2 - 398s_1 - 5435 = 0$$

$$398s_2 - 294s_1 - 656 = 0$$

$$-5435s_2 + 656s_1 + 10399 = [\epsilon\epsilon]$$

are calculated, from which we find

$$s_2 = 1.960557, \qquad s_1 = 0.422795, \qquad [\epsilon\epsilon] = 20.7.$$

As the mean value of an equation, we obtain

$$m = \left(\frac{20.7}{15}\right)^{1/2} = 1.175,$$

and as the mean error of a difference, $m_\Delta = 0.526$. From the reciprocal equation we get the values

$$u_{1,2} = 0.039329 \pm 0.999222i; \qquad u_{3,4} = -0.250726 \pm 0.968058i.$$

For the minimum values, we get

$$7\alpha = 87°.746; \qquad 7\beta = 104°.554.$$

Therefore, in sufficient agreement with the values for $h = 10$, we have

$$\alpha = 12°.53; \qquad \beta = 14°.94,$$

while all the other values are different. A determination of the coefficients A and B is of no further interest here.

NOTES

1. Willers, *Numerische Integration* (Berlin, 1923), pp. 79-82; Runge-König, *Numerisches Rechnen* (Berlin, 1924), pp. 246-8.

2. *The Calculus of Observations*, 2nd ed. (1926), p. 349. References are also given on p. 360 ff. See also *Enzyklopädie d. math. Wiss.* I, A, 9a. Burkhardt, *Trigonometrische Interpolation*.

3. Cf. *Naturwissenschaften* 14 (1926), pp. 637-8.

CHAPTER SIX

APPROXIMATE INTEGRATION OF ORDINARY DIFFERENTIAL EQUATIONS

31. Graphical Methods.

1. With a differential equation of first order,

$$y' = f(x, y), \tag{1}$$

where $f(x, y)$ may be a single valued function of x and y, a slope is assigned to each point of a certain region of the xy plane. If the equation should be integrated, this means that we would seek curves $y = \varphi(x)$ which have at each point the prescribed slope. These curves are known as integral curves of the equation (1). From the infinitely many integral curves, we can pick out a certain particular solution of the differential equation, perhaps from the condition that the integral curve should pass through the point with coordinates x_0, y_0. If the function $f(x, y)$ is continuous in the region in question, then the integral curves are differentiable throughout the entire region. The differential equation can also be given in implicit form:

$$F(x, y, y') = 0. \tag{2}$$

If we should carry out the integration graphically, then we must provide as many points as possible with the slope assigned to them by the differential equation. We mark these slopes, or tangents, by short lines, and then draw in curves which have these prescribed slopes. It is to be observed that this problem is completely independent of the analytic formulation. The slopes assigned at the various points can just as well be given by observations as through calculation from a function. The drawing of the curves which have at each place the prescribed slope always corresponds to the integration of a differential equation of first order. In order to avoid an unnecessarily large number of such tangent lines (which would endanger the clarity of the drawing), we combine certain slopes or the slopes of certain lines. This makes no difficulties for analytical functions, and is also useful in the case of fields of tangents, given by obervations.

2. We connect the points of the same slope by curves, the so-called *isoclines*. In this way we carry out an interpolation, since we arrange all points through which the isocline passes according to the tangent

belonging to it, even those for which we have not made the calculation or the observation. If we have determined the slope at points which are not too far apart, we shall in general be able to assume a linear change of slope of the connecting line. Of course, for empirical functions, such as wind directions, this is often valid only for mean values in the immediate neighborhood.

The method of isoclines is also useful for functions given analytically. But we will then only use it if we have to draw a large number of integral curves, or if the isoclines are curves which can be easily constructed. It is simplest to assign a number to each isocline, and assign to it as slope a line of a *pencil of rays* which has the same number. The given data then consist of a family of curves and a pencil of lines. These are paired with one another according to the assigned numbers. This arrangement can

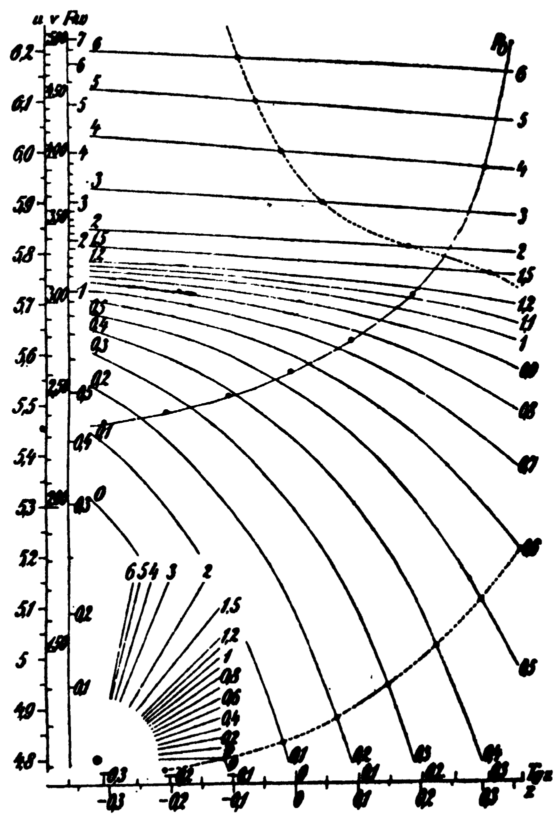


FIG. 118

also be made in the following way: we intersect each ray of the pencil of lines with the corresponding curve, as is done in Fig. 118, or we cut the

pencil of rays with a conveniently placed pair of coordinate axes. We then draw parallels to the x axis through the intersection point of each ray with the y axis up to the intersection with the corresponding isocline. The connecting line of these points is called the *directrix* by d'Ocagne.

If the differential equation is given in the form

$$F(x, y, y') = 0,$$

then the isocline to which is assigned the slope α has the equation

$$F(x, y, \alpha) = 0.$$

The directrix connects all points for which $\alpha = y$; it therefore has the equation

$$F(x, y, y) = 0. \tag{3}$$

We form the integral curves from pieces of straight lines whose intersection we place between two successive isoclines. We draw a continuous curve in this train of straight lines, which has the prescribed slope at the intersections with the isoclines. Naturally these curves give only a rather rough approximation, but they can be drawn very quickly, and can be improved by a process to be described in Art. 33^1.

3. *Example*: The Newton equation

$$m\dot{v} = -mg \sin \theta - W(v) \tag{3a}$$

holds for the motion of a body on which the attractive force mg of the earth and the air resistance W (which is a function of the velocity v) are acting. Here θ is the angle the tangent to the path makes with the horizontal. Now the line element of the path is $ds = \rho d\theta$, where ρ is the radius of curvature, and since $v = ds/dt$, we have $\rho d\theta = vdt$. For the centrifugal acceleration, we have $-g \cos \theta = v^2/\rho = v^2\, d\theta/ds = v d\theta/dt$. If we express v as a function of θ and θ as a function of t in the equation of motion, then, using the relations given above, we have

$$\begin{aligned} m \frac{dv}{d\theta} \cdot \frac{d\theta}{dt} + mg \sin \theta &= -\frac{m}{v}\frac{dv}{d\theta}\, g \cos \theta + mg \sin \theta \\ &= -\frac{mg}{v}\frac{d(v \cos \theta)}{d\theta} = -W(v) \end{aligned} \tag{3b}$$

i.e.,

$$\frac{d(v \cos \theta)}{d\theta} = + \frac{v \cdot W(v)}{mg}. \tag{3c}$$

To be able to handle this equation more conveniently, graphically or numerically, we introduce the new variables[2]

$$u = \ln v; \qquad z = \ln \operatorname{tg}\left(\frac{\pi}{4} + \frac{\theta}{2}\right), \qquad \text{i.e., } \tanh z = \sin \theta;$$

$$du = \frac{dv}{v}; \qquad dz = \frac{d\theta}{\cos \theta}.$$

If we express the weight of the projectile in kg, $G = mg$, then the equation of the projectile motion becomes

$$\frac{du}{dz} = \tanh z + \frac{W(e^u)}{mg} = \tanh z + \frac{W(e^u)}{G}.$$

The air resistance is

$$W = \frac{R^2 \pi \delta i}{1.22} K v^2$$

where R is the radius of the projectile in cm., i is a form factor for which we substitute 1, δ is the density of the air which, for not too high altitudes, is equal to 1.22, u is the projectile velocity in m/sec, and K is a quantity independent of v, which we obtain from tables.[3] For example, if we take $R = 5$ cm., $G = 12.21$ kg., then $R^2\pi\delta i/1.22G = 6.4$. Therefore we have

$$\frac{du}{dz} = \tanh z + 6.4K \cdot e^{2u} = \tanh z + F(u).$$

In Fig. 118, a scale for tanh z is laid along the z axis, and a scale first for v and then for $F(u)$ is laid along the u axis. From these two scales we easily construct the corresponding isoclines, to which are drawn the tangent lines having the same number. The intersection of the extended lines from the pencil with the corresponding isoclines is connected by a dotted curve. If we are only interested in the path of the projectile, then we will not draw in the entire path of the isoclines, but only that short part of them which is necessary for the construction. If, for example, we take the initial velocity to be $v_0 = 500$ m/sec., the initial angle as $\theta = 20°$, then $u_0 = 6.2146$, $z_0 = 0.3564$. The integral curve then is drawn out from this point P_0 by the use of isoclines. The path of the projectile is drawn in Fig. 119. In this we have

$$x = \int_{\theta_0}^{\theta} \frac{v^2}{g}\, d\theta \text{ (dashed line)}, \; y = -\int_{\theta_0}^{\theta} \frac{v^2}{g} \operatorname{tg} \theta \, d\theta \text{ (solid line)}$$

$$t = -\int_{\theta_0}^{\theta} \frac{v\,d\theta}{g \cos \theta} \text{ (dot-dash line).}$$

These integrations are carried out graphically in the drawing, and

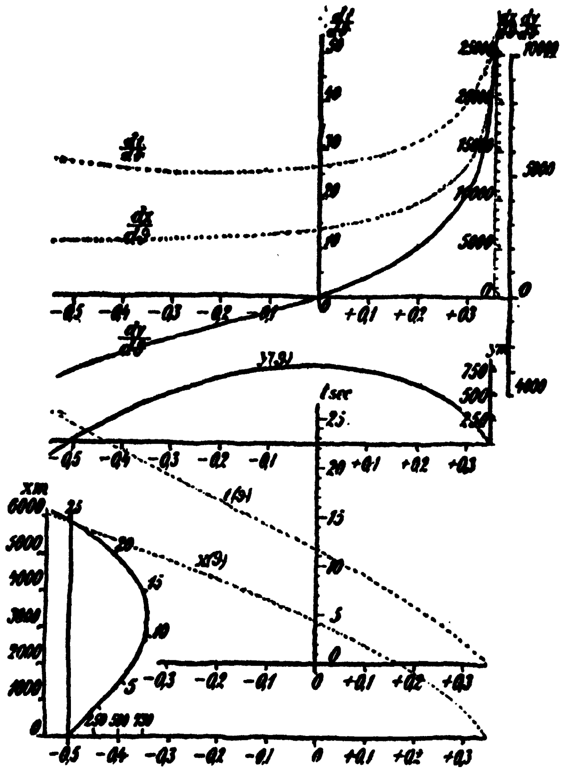


FIG. 119

the numbered curve is constructed at the lower left by taking the values x, y, t belonging to an appropriately chosen θ. This curve represents the path of the projectile. The numbers give the time after the firing at which the particular point of the path is reached.

4. Another possible way of collecting the slopes is the following: we take any curve K in the tangent field and extend the tangents assigned to the individual points of this curve. These straight lines cover a curve S, which is known as the *ray curve*. If we draw a ray curve for each curve used, then we need only draw the tangents to the assigned ray curve from any point of the first curve, to get the slope prescribed at this point. Naturally we choose as the initial curve only those curves which are

easy to draw, and we apply the method only to such equations which lead to ray curves which are easily drawn.[4]

The simplest method is to join the slopes of the lines $y = mx + n$. To find the equation of the ray curve belonging to such a straight line, we

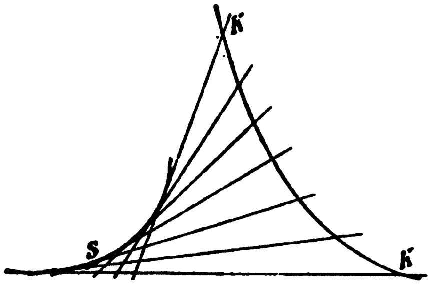


FIG. 120

start out from the differential equation (1), $y' = f(x, y)$. If we denote the running coordinates of the line determined by a tangent element of the point x, y by ξ, η, then the equation of this straight line is

$$\eta - y = f(x, y)(\xi - x). \tag{4}$$

The equation of the tangent line belonging to a neighboring point with the coordinates $x + \Delta x$, $y + \Delta y$ becomes

$$\begin{aligned} \eta - y - \Delta y &= f(x + \Delta x, y + \Delta y)(\xi - x - \Delta x) \\ &= (f(x, y) + f_x(x, y)\Delta x + f_y(x, y)\Delta y)(\xi - x) \\ &\qquad - f(x, y)\cdot\Delta x + \cdots . \end{aligned} \tag{4a}$$

By consideration of equation (4), we get, for the intersection of the two lines,

$$-\Delta y = (\xi - x)(f_x(x, y)\Delta x + f_y(x, y)\Delta y) - f(x, y)\Delta x + \cdots . \tag{5}$$

Since the points in question should lie on the line $y = mx + n$, then $\Delta y = m\Delta x$; therefore, upon division by Δx, we get

$$-m = (f_x(x, y) + mf_y(x, y))(\xi - x) - f(x, y) + \cdots , \tag{6}$$

where the omitted terms contain Δx and higher powers of Δx. If we let Δx approach zero, then ξ becomes a point of the envelope. From (6) it then follows that

$$\xi = x + \frac{f(x, y) - m}{f_x(x, y) + mf_y(x, y)}, \tag{7}$$

and if we substitute this in (4), we get for the ordinates belonging to the abscissa ξ of the point of the ray curve

$$\eta = y + \frac{f(x, y)(f(x, y) - m)}{f_x(x, y) + mf_y(x, y)}. \tag{8}$$

We substitute $y = mx + n$ in this equation, getting x as a parameter. If we want to have the equation of the ray curve in closed form, then we must eliminate x from (7) and (8). The construction becomes especially simple if the ray curve contracts into a point, the *ray point*. This is the case if ξ and η are independent of x and y individually, but are functions of $(y - mx)$. If the equation (1) has the form

$$y' = \frac{y - g(y - mx)}{x - h(y - mx)} = \frac{Z}{N}, \tag{9}$$

then

$$\begin{aligned} \frac{f - m}{f_x + mf_y} &= \frac{Z - mN}{N} \cdot \frac{N^2}{Nm - Z} = -N = -x + h(y - mx), \\ \frac{f(f - m)}{f_x + mf_y} &= -N \cdot \frac{Z}{N} = -Z = -y + g(y - mx). \end{aligned} \tag{10}$$

In fact, ξ and η are then only dependent on $y - mx = n$. Therefore a ray point belongs to each straight line.

If we choose parallels to the y axis as the straight lines, then $x = C$, and we get ray point according to (9), if the equation has the form

$$y' = \frac{y - g(x)}{x - h(x)}. \tag{10a}$$

This is a linear differential equation of first order $y' = A(x)y + B(x)$, where

$$A(x) = \frac{1}{x - h(x)}, \qquad B(x) = -\frac{g(x)}{x - h(x)} = -g(x)\cdot A(x). \tag{10b}$$

From (7), (8), and (10), it then follows[5] that

$$\xi - x = -1/A(x), \qquad \eta = -B(x)/A(x). \tag{10c}$$

In particular, if $A(x)$ is a constant, then the ray point is always at the same distance from the corresponding y-parallel. One of the integrators of Pascal (35.3) is based on this principle.

5. We could also join the slopes on arbitrarily chosen curves $\varphi(x, y) = c$. In this case, we would get $\Delta y = -\varphi_x\Delta x/\varphi_y + a(\Delta x)^2 + \cdots$. If we substitute this in (5), then the equation of the ray curve becomes

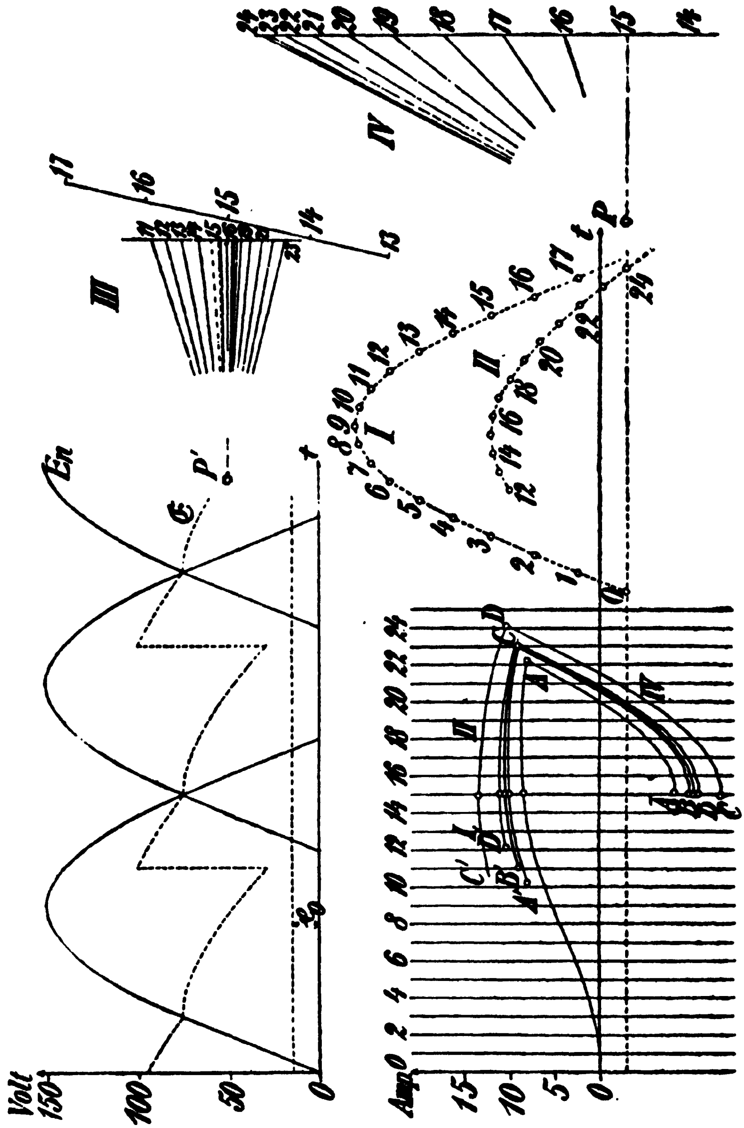


FIG. 121

$$(11)\qquad \xi = x + \frac{\varphi_y \cdot f(x, y) + \varphi_x}{\varphi_y \cdot f_x(x, y) - \varphi_x f_y(x, y)}, \quad \eta = y + \frac{f(x, y)(\varphi_y \cdot f(x, y) + \varphi_x)}{\varphi_y f_x(x, y) - \varphi_x f_y(x, y)}$$

if $\phi_y \neq 0$. The ray curve is reduced to a ray point in this case if the differential equation takes the form

(11a) $$y' = \frac{y - g(\varphi(x, y))}{x - h(\varphi(x, y))},$$

as can easily be confirmed by calculation. We can set up types of equations for which the joining of the slope on certain curves gives ray points.[6]

The application of ray curves is then useful only if these curves are easy to construct, or if simple tangent construction can be carried out without drawing the curve, as is the case for conic sections, for example.[7]

6. *Example*: The *polyphase mercury vapor rectifier*[8] works so that during one time interval only one anode delivers current. Then this anode delivers current in conjunction with the next one, then the next anode alone delivers current, etc. Therefore periods of single anode operation alternate with transition periods. For the single operation period, the current density J is determined from the equation

(I) $$J' = -\frac{R_1}{L_1}J + \frac{E_0 \sin \omega t - E_g}{L_1},$$

and for the transition period from

(II) $$I' = -\frac{R_2}{L_2}I + \frac{E_0 \sin \omega t + E_0 \sin(\omega t - \varphi) - 2E_g}{L_2},$$

where R_a and L_a or R_K and L_K are the resistance and inductance of the anode or of the direct current network respectively, and where

$$R_1 = R_a + R_K\,;\ L_1 = L_a + L_K\,;\ R_2 = R_a + 2R_K\,;\ L_2 = L_a + 2L_K\,.$$

E_0 and ω are the maximum potential and the angular frequency of the alternating current, φ is the phase angle difference of two successive currents, and E_g is the sum of the potential drop e_0 in the rectifier and the counter e.m.f. E_K of the network. Both of these may be regarded as constant.

The *end* t_1 *of the single anode period* is determined by the condition

(III) $$\begin{aligned} L_aJ' + R_aJ &= E_0 \sin \omega t - E_0 \sin(\omega t - \varphi) \\ &= -2E_0 \cos\left(\omega t - \frac{\varphi}{2}\right) \sin \frac{\varphi}{2}, \end{aligned}$$

and the *end of the transition period* by

(IV) $$j' = -\frac{R_a}{L_a}j - \frac{E_0 \sin \omega t - E_0 \sin(\omega t - \varphi)}{L_a},$$

where the curve for j is to be constructed from the point with the

abscissa t_1, i.e., the time at which the single operation ends, and with the ordinate $-J$, which is therefore equal and opposite to the ordinate of the current curve at this point, up to the intersection with the I curve. There the time is t_2 for the end of the transition period, and the cycle begins again. The value J_2 of the current at the time t_2 must, for steady operation, be equal to the initial value J_0 at the beginning of the single operation period. To get this, we start out from an arbitrary value $\bar{J}_0$, construct the curve for J and take the final value $\bar{J}_2$ as the new initial value $\bar{\bar{J}}_0$, etc., until we can no longer distinguish the initial and final values from each other, within the accuracy of the drawing.

All the equations (I), (II), (IV) are linear differential equations of first order with constant coefficients $A(x)$. Therefore the method described in the preceding section is suitable. A ray point is then assigned to each y-parallel, and we obtain

$$\text{(Ia)} \qquad \xi = t + \frac{L_1}{R_1}; \qquad \eta = \frac{E_0 \sin \omega t - E_g}{R_1},$$

$$\text{(IIa)} \qquad \xi = t + \frac{L_2}{R_2}; \qquad \eta = \frac{E_0 \sin \omega t + E_0 \sin(\omega t - \varphi) - 2E_g}{R_2},$$

$$\text{(IVa)} \qquad \xi = t + \frac{L_a}{R_a}; \qquad \eta = \frac{E_0[\sin(\omega t - \varphi) - \sin \omega t]}{R_a}.$$

The ray points for each equation then are at the same distances from the corresponding y-parallels, and the curves drawn through the ray points are sine curves.

The following should be observed with regard to the units of measurement: if the units are given by the lengths

$$1 \text{ sec} = \tau \text{ cm.}, 1 \text{ amp} = i \text{ cm.}, 1 \text{ volt} = \epsilon \text{ cm.}, 1 \text{ ohm.} = \rho \text{ cm.},$$

$$1 \text{ henry} = \lambda \text{ cm.},$$

then

$$\rho = \frac{\epsilon}{i}; \qquad \lambda = \frac{\epsilon\tau}{i}.$$

Therefore only three units of measurement can be chosen arbitrarily.

In Fig. 121, the construction is carried out for a three phase rectifier. In this case

$$\varphi = 2\pi/3; \qquad E_0 = 150\text{v}; \qquad R_a = 0; \qquad R_K = 5\Omega; \qquad E_g = 15\text{v};$$

$$L_a = 0.03\text{h}; \qquad L_K = 0.03\text{h}.$$

The frequency is to be taken as $c = 60/\text{sec}$. Then

$$R_1 = 5\Omega; \qquad R_2 = 10\Omega; \qquad L_1 = 0.06\text{h}; \qquad L_2 = 0.09\text{h}.$$

We take 9 cm. as the scale modulus for the half period (Fig. 121 is reduced to half this value), i.e.,

$$\tau = 1080 \text{ cm}, i = \frac{1}{4} \text{ cm}, \epsilon = \frac{1}{20} \text{ cm, then } \rho = \frac{1}{5} \text{ cm}, \lambda = 216 \text{ cm}.$$

If we denote the length in centimeters by primes, we have

$$E_0' = 7.5 \text{ cm}; E_g' = 0.75 \text{ cm}; R_K' = 1 \text{ cm}; L_a' = L_K' = 6.48 \text{ cm};$$

$$R_1' = 1 \text{ cm}; R_2' = 2 \text{ cm}; L_1' = 12.96 \text{ cm}; L_2' = 19.44 \text{ cm}.$$

The ray points are determined by

(I) $\xi' - t' = 12.96$ cm, $\eta' = 7.5 \sin(120\pi t - 0.75)$;

(II) $\xi' - t' = 9.72$ cm, $\eta' = 3.75 \sin\left(120\pi t - \frac{\pi}{3}\right) - 0.75$;

(IV) $\xi' - t' = \infty$, $\eta' = \infty$;

$$\frac{\eta'}{\xi' - t'} = \frac{7.5}{6.48} (3)^{1/2} \cos\left(\omega t - \frac{\pi}{3}\right) \approx 2 \cos\left(120\pi t - \frac{\pi}{3}\right).$$

In case (IV) therefore, we have no ray points, but a pencil of lines. We draw the y-parallel at a distance of 0.5 cm., and then the ray point (I) (dotted curve with open circles). Then we draw the integral curve from zero out and construct $L_a J^?$ graphically, to determine for which value of t this is equal to the ordinate of the curve $-12.98 \cos(\omega t - \pi/3)$ drawn on the upper right, under (III), for the values 13 to 17. If this is the case for the abscissa t_1, then this is the endpoint of the single operation period, marked by an open circle. From here on the integral curve is drawn by use of the ray point (II). Simultaneously, the integral curve of the equation (IV) is drawn from the point A with the coordinates t_1, $-J$, by use of the pencil of lines (IV). Both intersect in a point A with abscissa t_2. This is the endpoint of the transition period. This point is shifted $1080/3\times 60 = 6$ cm. to the left, to A'. If A' were to lie on the integral curve, then the desired current would then be found for the steady state. Since this is not the case, we carry out the construction of A' again and repeat it until the resultant point lies on the curve. After a second repetition, C coincides with B' within the limits of accuracy of the drawing.

For a check, the construction is carried through again from a point C', which is higher up on the drawing. Here also we get coincidence after a second repetition. In addition to the potential curve E_n, the anode potential $\mathfrak{E}$ is also plotted in the upper left portion, as a dotted line. This is calculated from the constructed current curve by the equation

$$E_n - L_1 J' - R_1 J = \mathfrak{E}_n \text{ for the single operation period;}$$
$$1/2(E_n + E_{n+1} - L_a J' - R_a J) = \mathfrak{E}_n \text{ for the transition period.}$$

NOTES

1. Details on the integral curves can be found in Willers, *Graphische Integration* (Berlin, 1920), Art. 12 and Art. 13.
2. Cranz and Rothe, *Art. Monatshefte* (1917), p. 197.
3. Cranz, *Lehrbuch der Ballistik* I (Berlin, 1925), p. 61.
4. Mehmke, *Leitfaden zum graphischen Rechnen* (Leipzig, 1917), p. 119.
5. Czuber, *Z. f. Math. u. Phys.* 44 (1899), p. 41.
6. A corresponding treatment for polar coordinates is found in Neuendorf, *Z. f. angew. Math. u. Mech.* 2 (1922), p. 131.
7. Willers, *Arch. f. Math. u. Phys.* III, 26 (1918), p. 96, and III, 27 (1918), p. 51.
8. Pflieger-Haertel, *Wissenschaftliche Veröffentlichungen aus dem Siemens Konzern*, III (1923), p. 61.

32. Numerical Methods.

1. In numerical integration, the problem is to *calculate additional points of the integral curve, starting from a point* x_0, y_0 *of a particular solution.* In order to find another point we must calculate the change k which y undergoes if we change x by h. We use the Taylor series for this calculation:

$$k = y_0' \cdot h + y_0'' \frac{h^2}{2!} + y_0''' \frac{h^3}{3!} + y_0^{(4)} \frac{h^4}{4!} \cdots, \tag{1}$$

where, if we start from the equation $y' = f(x, y)$,

$$\begin{aligned} y_0' &= f(x_0, y_0), \\ y_0'' &= f_x(x_0, y_0) + f(x_0, y_0) \cdot f_y(x_0, y_0), \end{aligned} \tag{1a}$$

etc. We soon obtain a very complicated expression. To write this down in simple fashion, we introduce a symbolic representation, using the so-called operators, and set

$$\begin{aligned} D^1(\varphi) &= \varphi_x + f \cdot \varphi_y, \\ D^2(\varphi) &= (\varphi_x + f\varphi_y)^2 = \varphi_{xx} + 2f\varphi_{xy} + f^2\varphi_{yy}, \end{aligned} \tag{1b}$$

etc. In order to be able to perform simple calculations with these operators, we need the following three rules, which may be easily derived:

$$D^1(\varphi + \psi) = D^1(\varphi) + D^1(\psi),$$

(1c)
$$D^1(\varphi \cdot \psi) = \varphi D^1(\psi) + \psi D^1(\varphi),$$

$$D^1(D^n(\varphi)) = D^{n+1}(\varphi) + nD^1(f)D^{n-1}(\varphi_y).$$

By use of these operators, the derivatives entering into the Taylor series may be written (if we omit the arguments):

$$y_0' = f,$$

$$y_0'' = D^1(f),$$

$$y_0''' = D^2(f) + f_yD^1(f),$$

(2)
$$y_0^{(4)} = D^3(f) + 2D^1(f)D^1(f_y) + f_yD^2f + f_y^2D^1(f) + D^1(f_y)D^1(f),$$

$$= D^3(f) + f_yD^2(f) + f_y^2D^1(f) + 3D^1(f)\cdot D^1(f_y),$$

$$y_0^{(5)} = D^4(f) + f_yD^3(f) + f_y^2D^2(f) + f_y^3D^1(f) + 4D^2(f)D^1(f_y),$$

$$+ 6D^1(f)D^2(f_y) + 7f_yD^1(f)D^1(f_y) + 3f_{yy}(D^1(f))^2.$$

These expressions are to be substituted in the series (1). If we continue the series far enough, we can calculate k from it with sufficient accuracy, provided that the series converges rapidly. But the calculation will, in most cases, be extremely involved, because the number of terms of the individual differential quotients increases greatly with increasing order. In this way it happens that for a function $f(x, y)$ given in tables, the calculation of the higher derivatives becomes inaccurate.

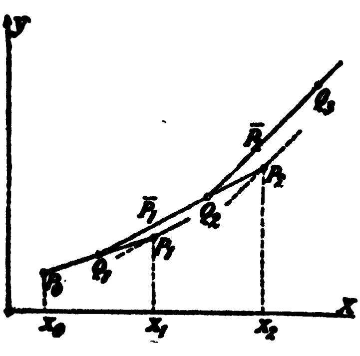


FIG. 122

2. Therefore, we must follow another method; in this, assuming satisfactory convergence of the series (1), we call that expression *the best approximate solution which gives, exactly, as many terms of this series as possible,* beginning with the first term. In many cases, as a first, rather rough approximation, we approximate the curve by use of the tangent segments, i.e., if we start out from the point x_0, y_0, we first set $k_0 = f(x_0, y_0)h_0$. A second point is determined by this. Furthermore we de-

termine the change of y_1 by means of the equation $k_1 = f(x_1, y_1)h_1$, etc. The error term in this case is proportional to h^2. The process is known as the *Cauchy difference method.*[1] In this way we can sketch a first approximation for the integral curve. Here we do not at first use the slope $f(x_1, y_1)$ at the point $P_1(x_1, y_1)$. Instead we join a segment of slope $f(x_1, y_1)$ at the point Q_1 (with abscissa $(x_0 + x_1)/2$) to the line train, and continue this line segment to the point $P_2(x_2, y_2)$. We then calculate the slope again, and join a line segment of this slope at the point Q_2 of the polygon which has the abscissa $(x_1 + x_2)/2$, etc. The approximation curve becomes still more accurate if we again determine the slope $f(x_1, \bar{y}_1)$ at the point P_1, in which the straight line joined at Q_1 cuts the y-parallel through x_1. We then join a line segment of this slope at the point Q_1. This method is continued until the slope no longer changes.

We can get about the same accuracy numerically, as was obtained in the graphical methods described above, with a method given by *Duffing.*[2] For an integration formula, he uses

$$k = y_1 - y_0 = \frac{h}{6}[4f(x_0, y_0) + hD^1(f(x_0, y_0)) + 2f(x_1, y_1)].$$

This is an equation in which y_1 enters as the unknown. We can therefore calculate the ordinate value y_1 of the integral curve, corresponding to the abscissa value $x_1 = x_0 + h$ by a solution of the equation for y_1. This method requires a great deal of calculation, since the equation for y_1 will, in general, be solvable only by approximation methods. It should be added that we have to form the first derivative of the function $f(x, y)$. To estimate the approximation, we can develop the last function in the brackets and get

$$\kappa = hf + \frac{h^2}{2} D^1(f) + \frac{h^3}{6} [D^2(f) + f_y D^1(f)]$$

$$+ \frac{h^4}{18} [D^3(f) + f_y D^2(f) + f_y^2 D^1(f) + 3D^1(f)D^1(f_y)] + \cdots .$$

The solution therefore agrees with the exact value (1) up to (and including) the term with h^3 as factor. This method can be extended to differential equations of second order. This is used by Funk[3] in the following way. He starts out from the function values and derivatives for the abscissas x_0 and $x_0 + h$, and calculates them for $x_0 + 2h$. In this way the terms of sixth order will be given correctly.

3. The formulas as were first given by *Runge*[4] are much easier to calculate. These formulas, which have the same accuracy as the formulas of Duffing, have been so improved by *Heun*[5] and *Kutta*[6] that we can represent

with 4 function values even the terms which contain h^4 as a factor. The chief advantage of these formulas is that they are obtained with the function values alone, so that we avoid, for the most part, the detailed calculation of the derivatives. The formulas represent a generalization to differential equations of the mean value methods described in Art. 16. We seek to represent the increment k which is added to the ordinate y_0 of the integral curve when the abscissa x_0 changes by h, as the mean value of the individual function values $f(x^{(n)}, y^{(n)})$ lying in the observed interval. We therefore write

$$k = \sum A_n f(x^{(n)}, y^{(n)}) = \sum A_n k^{(n)}.$$

Here the ordinate difference $y^{(m)} - y_0$ is denoted by $k^{(m)}$, and, for example, with the use of four function values, this will become

$$\begin{aligned} k' &= f(x_0\,, y_0)h, \\ k'' &= f(x_0 + \alpha h, y_0 + \alpha' k')h, \\ k''' &= f(x_0 + \beta h, y_0 + \beta' k' + \beta'' k'')h, \\ k^{(4)} &= f(x_0 + \gamma h, y_0 + \gamma' k' + \gamma'' k'' + \gamma''' k''')h, \end{aligned}$$

where the A and the α', β', β'' $\cdots$ are constants still to be chosen. The requirement that the approximation of the terms up to those multiplied by h^4 should agree with the development (1) gives 8 condition equations. We have 10 constants at our disposal for the fulfillment of these equations.[7] There are only 10 constants because the conditions $\alpha = \alpha'$, $\beta = \beta' + \beta''$, $\gamma = \gamma' + \gamma'' + \gamma'''$ must be satisfied. Therefore we have a *doubly infinite manifold of solutions* for the use of four function values.

4. One of these, given by Kutta, is quite important. Because of its symmetric construction, it is easily set up, and because of its small coefficients, it is especially useful for calculation. We have

$$\text{(3)} \qquad \begin{aligned} k' &= f(x_0\,, y_0)\cdot h, \\ k'' &= f\left(x_0 + \frac{1}{2}h, y_0 + \frac{1}{2}k'\right)\cdot h, \\ k''' &= f\left(x_0 + \frac{1}{2}h, y_0 + \frac{1}{2}k''\right)\cdot h, \\ k^{(4)} &= f(x_0 + h, y_0 + k''')\cdot h, \end{aligned} \qquad \begin{aligned} p &= \frac{1}{2}(k' + k^{(4)}) \\ q &= \frac{1}{2}(k'' + k''') \\ k &= \frac{1}{3}(p + 2q). \end{aligned}$$

If y does not appear on the right side of the differential equation, we then

have a simple quadrature, and this formula goes over into Simpson's rule. In this case then, the terms with the factor h^4 are still represented exactly. To prove that this also holds in general, and to determine the deviation of the term with the factor h^5, we develop the above equations in series:

$$k' = f \cdot h$$

$$k'' = f \cdot h + D^1(f)\frac{h^2}{2} + D^2(f)\frac{h^3}{8} + D^3(f)\frac{h^4}{48} + D^4(f)\frac{h^5}{384}\cdots$$

$$\begin{aligned} k''' = {} & f \cdot h + D^1(f)\frac{h^2}{2} + D^2(f)\frac{h^3}{8} + D^3(f)\frac{h^4}{48} + D^4(f)\frac{h^5}{384}\cdots \\ & + f_v D^1(f)\frac{h^3}{4} + f_v D^2(f)\frac{h^4}{16} + f_v D^3(f)\frac{h^5}{96}\cdots \\ & + D^1(f)D^1(f_v)\frac{h^4}{8} + [D^1(f_v)D^2(f) \\ & + f_{vv}(D^1(f))^2]\frac{h^5}{32} \\ & + D^1(f)D^2(f_v)\cdot\frac{h^5}{32} \end{aligned}$$

(3a)

$$\begin{aligned} k^{(4)} = {} & f \cdot h + D^1(f)h^2 + D^2(f)\frac{h^3}{2} + D^3(f)\frac{h^4}{6} + D^4(f)\frac{h^5}{24} \\ & + f_v D^1(f)\frac{h^3}{2} + f_v D^2(f)\frac{h^4}{8} + f_v D^3(f)\frac{h^5}{48} \\ & + f_v^2 D^1(f)\frac{h^4}{4} + (f_v^2 D^2(f) \\ & + 2f_v D^1(f)D^1(f_v))\frac{h^5}{16} \\ & + D^1(f)D^1(f_v)\frac{h^4}{2} + [D^1(f_v)D^2(f) \\ & + 2D^1(f)\cdot f_v D^1(f_v) \end{aligned}$$

$$+ f_{yy}(D^1(f))^2]\,\frac{h^5}{8}$$

$$+ D^1(f)D^2(f_y)\,\frac{h^5}{4}.$$

From these we get

$$
\begin{aligned}
k &= f\cdot h + D^1(f)\,\frac{h^2}{2} + [D^2(f) + f_yD^1(f)]\,\frac{h^3}{6}\\
&\quad + [D^3(f) + f_yD^2(f) + f_y^2D^1f + 3D^1(f_y)D^1(f)]\,\frac{h^4}{24}\\
&\quad + \left[\frac{25}{24}\,D^4(f) + \frac{5}{6}\,f_yD^3(f) + \frac{5}{4}\,f_y^2D^2(f) + \frac{15}{2}\,f_yD^1(f)D^1(f_y)\right.\\
&\quad \left. + \frac{15}{4}\,D^1(f_y)D^2(f) + \frac{15}{4}\,f_{yy}(D^1(f))^2 + \frac{25}{4}\,D^1(f)D^2(f_y)\right]\frac{h^5}{120}\cdots.
\end{aligned} \tag{4}
$$

Therefore these terms agree with those given in (2) up to those with the factor h^4, while the difference of the factors of h^5 gives the error of the approximation formula.

5. It is not possible to bring terms of fifth order into agreement with four function values, since 16 condition equations would have to be satisfied in this case. This is not possible with five function values either, since then only 15 constants are available.[8] Kutta gives formulas which do bring the factor of h^5 into agreement, using six function values. However, we can bring some of the terms of fifth order into agreement using four function values. For example, Runge[9] gives a formula in which the terms which do not contain f_y agree in the factor of h^5 as well as that of h^6. Such a formula then gives a better approximation if $f(x, y)$ changes only slightly with y, if then f_y is small. However, this formula is not so convenient for calculation as those given above.

If the function $f(x, y)$ becomes large in the course of the calculation, we shall get more accurate results if we exchange the variables. That is, instead of the equation $dy/dx = f(x, y)$, we use the equation $dx/dy = 1/f(x, y)$. We carry out this exchange, which we can make for each step of the calculation, whenever the absolute value of $f(x, y)$ is much larger than 1.

6. *The calculations here* must naturally be done schematically, and the following scheme is appropriate:

x	y	$f(x, y)$	$h \cdot f(x, y)$		y_0
x_0	y_0	$f(x_0\,, y_0)$	k'	k'	
$x_0 + \frac{h}{2}$	$y_0 + \frac{k'}{2}$	$f\left(x_0 + \frac{h}{2}, y_0 + \frac{k'}{2}\right)$	k''	$2k''$	
$x_0 + \frac{h}{2}$	$y_0 + \frac{k''}{2}$	$f\left(x_0 + \frac{h}{2}, y_0 + \frac{k''}{2}\right)$	k'''	$2k'''$	
$x_0 + h$	$y_0 + k'''$	$f(x_0 + h, y_0 + k''')$	$k^{(4)}$	$k^{(4)}$	k
				$6k$	y_1
x_1	y_1	$f(x_1\,, y_1)$	k'	k'	
$x_1 + \frac{h}{2}$	$y_1 + \frac{k'}{2}$	$f\left(x_1 + \frac{h}{2}, y_1 + \frac{k'}{2}\right)$	k''	$2k''$	
......	....	....	...	....	

In the scheme, calculations are carried out in horizontal rows, and the value of $k^{(m)}$ calculated each time is used in the next row.

To have a measure for the accuracy of the calculated values, we do not calculate the difference of the factors of h^5 given in (2) and (4), since the work of calculation would be too great. It is much easier to carry out the same calculation once again with the doubled interval $2h$. In the first case, let y have the increments k_1 and k_2 successively, in the second the increment $\bar{k}$. If the error of increment in the first case is $\Delta_1 = Ah^{n+1}$, then it becomes $\Delta_2 = A(2h)^{n+1}$ in the second case, if we neglect the terms with higher powers of h. We now have the identity

$$2\Delta_1 = \frac{\Delta_2 - 2\Delta_1}{2^n - 1} \approx \frac{\bar{k} - (k_1 + k_2)}{2^n - 1};$$

we can then say: the error after a double increment of h is about $1/(2^n - 1)$ of the difference of the two results, i.e., about $1/15$ of this difference in this case. This allows an approximate estimate of the magnitude of the error, which is, in general, completely adequate.[10]

7. *Example*: We again use the *projectile problem*, already considered graphically in 31.3. The problem is to integrate the equation

$$\frac{du}{dz} = \tanh z + 6.4Ke^{2u} = \tanh z + F(u).$$

In the case considered in Fig. 118, $\tanh z + 6.4ke^{2u}$ is larger than 1. We therefore begin with the equation

$$\frac{dz}{du} = \frac{1}{\tanh z + F(u)}$$

with the initial values $u_0 = 6.2146$, $z_0 = 0.3564$. A column is inserted after u for $v = e^u$. This is needed for the determination of $F(u)$. For the increment of u, we take $\Delta u = -0.25$.

u	v	z	$\tanh z$	$F(u)$	$f(u,z)$	$h/f(u,z)$	$[k]$	
6.2146	500	0.3564	0.3420	6.397	6.739	−0.03710	−0.03710	0.3564
6.0896	441.2	0.3378	0.3255	4.943	5.269	−0.04745	−0.09490	
6.0896	441.2	0.3327	0.3209	4.943	5.269	−0.04749	−0.09498	
5.9646	389.4	0.3089	0.2994	3.676	3.975	−0.06289	−0.06289	−0.04831
							−0.28960	+0.3081
5.9646	389.4	0.3081	0.2987	3.676	3.975	−0.06289	−0.06289	
5.8396	343.6	0.2767	0.2699	2.239	2.509	−0.09964	−0.19928	
5.8396	343.6	0.2583	0.2527	2.239	2.492	−0.10032	−0.20064	
5.7146	303.3	0.2078	0.2049	0.943	1.148	−0.21775	−0.21775	−0.1134
							−0.68056	0.1947

From here on, $f(u, z)$ is smaller than 1, so that we now return to the original equation. We take $\Delta z = -0.1$ as the interval width. The scheme then becomes

z	u	v	$\tanh z$	$F(u)$	$f(u.z)$	$hf(u,z)$	$[k]$	
0.1947	5.7146	303.3	0.1923	0.9432	1.1355	−0.11355	−0.11355	5.7146
0.1447	5.6578	286.5	0.1437	0.7397	0.8834	−0.08834	−0.17668	
0.1447	5.6704	290.2	0.1437	0.7767	0.9204	−0.09204	−0.18408	
0.0947	5.6226	276.6	0.0942	0.6576	0.7518	−0.07518	−0.07518	−0.0916
							−0.54949	5.6230

etc. We get

z	0.0947	−0.0053	−0.1053	−0.2053	−0.3053	−0.4053	−0.5053
u	5.6230	5.5589	5.5130	5.4816	5.4624	5.4537	5.4539

The numbers are given here up to the point that the velocity again begins to increase, because of the descent of the projectile. These calculated points are plotted in Fig. 118 by small circles. These naturally deviate somewhat from the curve obtained graphically. To get some idea of the accuracy, we carry out the first part of the calculation with the doubled interval width:

u	v	z	$\tanh z$	$F(u)$	$f(u,z)$	$h{:}f(u,x)$	$[k]$	
6.2146	500	0.3564	0.3420	6.397	6.739	−0.0742	−0.0742	0.3564
5.9646	389.4	0.3193	0.3089	3.676	3.985	−0.1255	−0.2510	
5.9646	389.4	0.2937	0.2855	3.676	3.962	−0.1262	−0.2524	
5.7146	303.3	0.2302	0.2262	0.9432	1.1694	−0.4276	−0.4276	−0.1675
							−1.0052	0.1889

The error of the value of z at the end of the second interval will therefore amount to about four units of the last decimal place, i.e., about 2%.

8. A *very useful* method, which is based *on the use of the approximation formulas for integrals derived in the difference calculations of Art. 12,* has been developed by Adams.[11] In this method we use the Newton formula for the difference of the terms on the increasing diagonal [12(1)]:

$$\int_{x_0}^{x_0+h} f(x)\,dx = h\Big[y_0 + \frac{1}{2}\Delta^1_{-1/2} + \frac{5}{12}\Delta^2_{-1} + \frac{3}{8}\Delta^3_{-3/2}$$

$$+ \frac{251}{720}\Delta^4_{-2} + \frac{95}{288}\Delta^5_{-5/2} \tag{5}$$

$$+ \frac{19087}{60480}\Delta^6_{-3} + \frac{5257}{17280}\Delta^7_{-7/2}\cdots\Big] + R_{n+1},$$

where, for $-nh \leqq \tau \leqq +h$ and continuous functions, $f^{(n+1)}(t)$,

$$R_{n+1} = h^{n+2}f^{(n+1)}(\tau)\int_0^1 \binom{t+n}{n+1}dt, \tag{6}$$

by 12(3). If the given equation is $y' = f(x, y)$ and if the values of y, and consequently those of y' are known for a series of equidistant abscissa values $x_n = x_0 + nh$, then we can build up a difference scheme from the known values of y':

$$\begin{array}{ccccccccc}
y'_0 & & & & & & & & \\
 & \Delta^1_{1/2} & & & & & & & \\
y'_1 & & \Delta^2_1 & & & & & & \\
 & \Delta^1_{3/2} & & \Delta^3_{3/2} & & & & & \\
y'_2 & & \Delta^2_2 & & \Delta^4_2 \cdots & & & & \\
 & \Delta^1_{5/2} & & \Delta^3_{5/2} & & & & & \\
y'_3 & & \Delta^2_3 & & \cdot\cdot & & & & \\
 & \Delta^1_{7/2} & & \cdot\cdot & & & & & \\
y'_4 & & \cdot\cdot & & & & & & \\
 & \cdot\cdot & & & & & & & \\
\cdot\cdot & & & & & & & &
\end{array} \tag{6a}$$

Also, we can usĕ the formula (5) on this scheme. This formula becomes

$$
\begin{aligned}
(7)\qquad y_{n+1} - y_n = \int_{x_n}^{x_n+h} y'\,dx = h\Big(y'_n &+ \frac{1}{2}\Delta^1_{(2n-1)/2} + \frac{5}{12}\Delta^2_{n-1} \\
&+ \frac{3}{8}\Delta^3_{(2n-3)/2} + \frac{251}{720}\Delta^4_{n-2} + \frac{95}{288}\Delta^5_{(2n-5)/2}\cdots\Big).
\end{aligned}
$$

For example, if we have so chosen the intervals that we can neglect the differences of fifth order within the desired accuracy, then we need only five ordinates or four derivatives for the formulation of the initial difference scheme, and can calculate y_5 by means of the formula

$$
(8)\qquad y_5 = y_4 + h\Big(y'_4 + \frac{1}{2}\Delta^1_{7/2} + \frac{5}{12}\Delta^2_3 + \frac{3}{8}\Delta^3_{5/2} + \frac{251}{720}\Delta^4_2\Big).
$$

The equation $y'_5 = f(x_5, y_5)$ gives the next value of the first column of the difference scheme, so that we can complete it about a rising diagonal. With these new differences we find y_6 by (8)($n = 5$), then y'_6, etc. If we have the beginning of the scheme, then the further calculation can be carried out mechanically. If we assume the differences of a certain order are constant, then we can also work with the other formulas in Art. 12, and we can progress more rapidly. This can occasionally be useful for the rapid calculation of a first approximation, which we then improve by the methods of the following section.[12]

9. We now have several possibilities for the *calculation of the first value of the scheme*. For example, we can determine the value by means of the formulas of Runge-Kutta described above. But we can also calculate the coefficients of the Taylor series, given in (1), directly. The higher derivatives appearing there as coefficients can best be calculated directly from the equation; we then do not need to form the detailed values (2).

Example: We take the equation

$$
(8a)\qquad y' = (xy - 1)^{1/2}.
$$

From this equation, which we write in a somewhat different form, we find by repeated differentiation:

$$
(8b)\qquad
\begin{aligned}
y'^2 - xy + 1 &= 0 \\
2y'y'' - y - xy' &= 0 \\
2y'y''' + 2(y'')^2 - 2y' - xy'' &= 0 \\
2y'y^{(4)} + 6y''y''' - 3y'' - xy''' &= 0
\end{aligned}
$$

$$2y'y^{(5)} + 8y''y^{(4)} + 6(y''')^2 - 4y''' - xy^{(4)} = 0$$

$$2y'y^{(6)} + 10y''y^{(5)} + 20y'''y^{(4)} - 5y^{(4)} + xy^{(5)} = 0$$

$$2y'y^{(7)} + 12y''y^{(6)} + 30y'''y^{(5)} + 20(y^{(4)})^2 - 6y^{(5)} - xy^{(6)} = 0$$

$$2y'y^{(8)} + 14y''y^{(7)} + 42y'''y^{(6)} + 70y^{(4)}y^{(5)} - 7y^{(6)} - xy^{(7)} = 0.$$

If we take as the initial value $x = 1$, $y = 2$, then, according to the sign which we give the root, this becomes $y' = +1$ or $y' = -1$; if we calculate the higher derivatives for these values from the above equations, then we get, for example, for the negative value of the root,

$$\text{(8c)} \quad \begin{aligned} y &= 2 - h - \frac{1}{4}h^2 + \frac{1}{4}h^3 - \frac{3}{32}h^4 + \frac{5}{64}h^5 - \frac{5}{64}h^6 \\ &\quad + \frac{153}{1792}h^7 - \frac{2869}{28672}h^8 \cdots . \end{aligned}$$

We take $h = 0.02$ as the interval width. The error in one interval then becomes, by (6),

$$\text{(9)} \quad \begin{aligned} |R| &= h^6 \, |f^{(5)}(\tau)| \int_0^1 \frac{(t+4)(t+3)(t+2)(t+1)t}{5!}\,dt, \\ &= (0.02)^6 \cdot \frac{475}{12} \cdot \frac{f^{(5)}(\tau)}{5!}. \end{aligned}$$

if we again consider the four differences. For $h = 0$, $f^{(5)}(\tau) < 10$ and then decreases, as we can obtain from the above series. Then

$$\text{(9a)} \qquad |R| \approx \frac{950}{288}(0.02)^6 < 211 \cdot 10^{-12}.$$

Therefore, we will be able to carry the calculations to about 8 decimal places. From the above series, we first calculate the values y_0 to y_4, then, by means of the equation $y' = (xy - 1)^{1/2}$ we compute the values y_0' to y_4' from these. These values are entered in the third column. In the fourth column are the values hy', from which the difference scheme is then formed. The calculation is carried out with seven place logarithms. The irregularities in the fourth difference are reduced to rounding off errors.

x	y	y'	$h \cdot y'$	Δ^1	Δ^2	Δ^3	Δ^4
1.00	2	−1	−0.02000000				
				−19046			
1.02	1.97990199	−1.0097029	−0.02019406		+1166		
				−18240		−33	
1.04	1.95961577	−1.0188230	−0.02037646		+1133		+5
				−17107		−28	
1.06	1.93915284	−1.0273763	−0.02054753		+1105		0
				−16002		−28	
1.08	1.91852440	−1.0353774	−0.02070755		+1077		+2
				−14925		−26	
1.10	1.89774135	−1.0428401	−0.02085680		+1051		+2
				−13874		−24	
1.12	1.87681430	−1.0497771	−0.02099554		+1027		+3
				−12847		−21	
1.14	1.85575368	−1.0562004	−0.02112401		+1006		0
				−11841		−21	
1.16	1.83456963	−1.0621209	−0.02124242		+ 985		0
				−10856		−21	
1.18	1.81327212	−1.0675491	−0.02135098		+ 964		
				− 9892			
1.20	1.79187088	−1.0724948	−0.02144990				
1.22	1.77037546						

Another possibility for getting the initial values of the difference scheme is the method of iteration to be described in the following article.

NOTES

1. See Hamel, *Elementare Mechanik*, 2nd ed. (Berlin, 1922), p. 115, No. 71, 2.
2. Duffing, *Forschungarbeiten aus dem Gebiet des Ingerieurwesens.*. No. 224 (1920).
3. Funk, *Z. f. angew. Math. u. Mech.* 7 (1927), p. 410.
4. Runge, *Math. Annalen* 46 (1895), p. 167.
5. Heun, *Z. f. Math. u. Phys.* 45 (1900), p. 23.
6. Kutta, *Z. f. Math. u. Phys.* 46 (1901), p. 435.
7. Runge-König, *Numerisches Rechnen* (Berlin, 1924), p. 290 ff.
8. The second group of formulas there contains an error which has been corrected by Nyström, *Acta societates scientiarum fennicae* 50 (1925), No. 13, p. 5.
9. Runge-König, *op. cit.*, pp. 299–300.
10. Runge has investigated the effect of an error of the original data on the further calculations. *Math. Annalen* 44 (1894), p. 437; *Göttinger Nachrichten* (1905), p. 252.
11. Bashforth and Adams, *An Attempt to Test the Theories of Capillary Action* (Cambridge, 1883), p. 18.
12. Lindow, *Numerische Infinitesimalrechnung* (Berlin, 1928), Ch. IV.

33. Method of Iteration.

1. For the usefulness of the methods described in the two preceding articles, it is essential that, if the accuracy of the solutions found by these

methods should not be sufficient, we can improve them graphically as well as numerically, as is shown in the simplest case in 32.3. It is also necessary in almost all cases which are of importance for practical use, that we can approach *the actual solution stepwise in particular intervals*. In this way, the degree of the approximation in graphical methods is limited only by the accuracy of the drawing, and in numerical methods only by the number of places carried out. This approximation process has been used by Schwarz, Picard and Lindelöff[1] for the proof of the existence of the solution, and was then applied by Runge, Cotton and others to the actual construction[2] or calculation[3] of the solution.

If we desire the solution of the equation

$$(1) \qquad y' = f(x, y)$$

which passes through the point x_0, y_0, we write this solution in the form

$$(2) \qquad y = y_0 + \int_{x_0}^{x} f(x, y)\, dx.$$

If we have any approximate solution $y_{(1)}$ of the equation (1), then, if this is substituted in the right side of (2) and the integral is evaluated graphically, according to the methods of Art. 14, or numerically, according to the methods of Art. 12, a new approximate solution $y_{(2)}$ is given on the left side:

$$(2a) \qquad y_{(2)}(x) = y_0 + \int_{x_0}^{x} f(x, y_{(1)})\, dx.$$

If we process with $y_{(2)}$ exactly as with $y_{(1)}$, with the new approximation $y_{(3)}$ in the same way, etc., then we finally obtain an $(n + 1)$st approximation

$$(3) \qquad y_{(n+1)} = y_0 + \int_{x_0}^{x} f(x, y_{(n)})\, dx.$$

To investigate the *convergence of the process*, we form the difference of the two equations (2) and (3):

$$(3a) \qquad y - y_{(n+1)} = \int_{x_0}^{x} [f(x, y) - f(x, y_{(n)})]\, dx.$$

If we denote the difference between the actual solution y and the approximation $y_{(n)}$ with $\epsilon_n = y - y_{(n)}$, then the equation

$$(4) \qquad \epsilon_{n+1}(x) = \int_{x_0}^{x} \frac{f(x, y) - f(x, y_{(n)})}{y - y_{(n)}}\, \epsilon_n(x)\, dx$$

is obtained. Since we are dealing with small deviations between the actual solution and the approximate solution, we have, approximately,

(4a) $$\frac{f(x, y) - f(x, y_{(n)})}{y - y_{(n)}} \approx \frac{\partial f}{\partial y}.$$

If we now assume that this value does not exceed a finite maximum M in the interval under consideration (the choice of the coordinate system can always be provided), and if we further assume that $| \epsilon_n |$ is the maximum absolute value of $\epsilon_n(x)$ in this interval, then it follows from the equation (4), by use of the mean value theorem of integral calculus, that

(4b) $$| \epsilon_{n+1} | \leqq M | \epsilon_n | | x - x_0 |.$$

Since M is a finite number, we can always choose the size of the integration interval so that $M | x - x_0 | = \kappa < 1$. In this case then,

(5) $$| \epsilon_{n+1} | \leqq \kappa^n | \epsilon_1 |,$$

and for sufficiently small values of κ soon differs only very slightly from zero.

In practice, we first obtain an approximation in a sufficiently small interval. This approximation does not need to be especially good; if we only choose the interval, and consequently κ sufficiently small, we shall, through repeated use of the method characterized by equation (3) soon approach the correct solution in this small interval, provided that M is not too large. If one approximation no longer differs from the preceding one, we then can assume that it agrees with the desired solution within the possible accuracy. We then proceed in a similar fashion in a neighboring small interval, starting from the endpoint of the last approximation, etc. In this way we can find approximations over larger intervals, even if the process only converges in a small interval. We can therefore apply this method directly to integration. In this case, we can start out fron any approximation value. In particular, we can obtain the initial values for the difference scheme in this way, if we want to use to the method described at the end of the preceding article.

2. The application of the method can best be understood by a. example. We take the example for which we calculated a small par of the solution in 32.9. We shall first find an approximate solutio graphically, by the method of the ray curve, improve this graphicall, and then further improve the resultant approximation by the numerical method. We then seek the solution of the equation

(5a) $$y' = (xy - 1)^{1/2}$$

which goes through the point $x_0 = 1$, $y_0 = 2$, for which the root is to be taken with negative sign. A first approximation is obtained by the Cauchy method. The determination of the slope is especially

simple here with the use of the method of ray curves explained in Art. 31. The curves are parabolas in this case.

We can find a still better approximation by the method given in 32.2. This first approximation is not drawn in Fig. 123 in the first interval, and in the second interval (between $x = 2$ and $x = 2.8$) it is the dashed line.

With the construction of the first approximation by tangent seg-

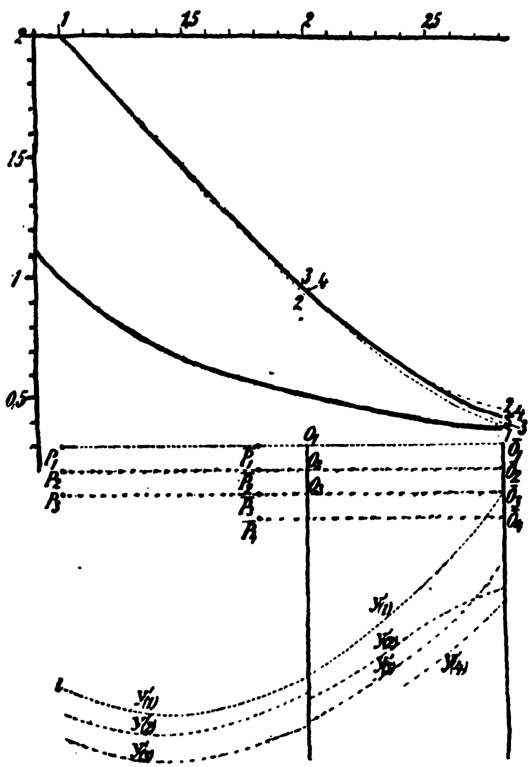


Fig. 123

ments we also have the corresponding slopes and can plot points of the curve $y'_{(1)} = f(x_1 y_{(1)})$ in suitably chosen units. We then draw a smooth curve through these points. This line is dotted in Fig. 123 and runs from the lower left to the dotted x axis through P_1. The graphical approximation in this figure is first carried through for the interval from 1 to 2. We now draw the integral curve to this dotted curve through the initial point A with the coordinates $x_0 = 1$, $y_0 = 2$. This is done by the method of Art. 14. In this case we must so choose

the integration base $P_1 0_1$ that we have the same scale modulus in the direction of the y axis with which the first approximation was drawn. This gives the dot-dash curve 2. We take the pair of values $x, y_{(2)}$ from this curve, substitute them in $y' = (xy - 1)^{1/2}$ and calculate the values of $y'_{(2)}$ which belong to the given x. These values are also plotted on the lower left. We draw the dot-dash curve $y'_{(2)}$ through them. We integrate this curve with the base $P_2 0_2$, get the new integral curve 3, etc. The next step gives a new integral curve 4 (not drawn) which differs only slightly from the approximation value previously obtained.

For the next interval, from 2 to 2.8, we draw the dashed approximation line 1, with the curve $y'_{(1)}$, from the endpoint of the last approximation. We again apply the method described above. We can see that the convergence is much less rapid in this interval, especially on the different paths of the curves $y'_{(1)}$ and $y'_{(2)}$. We have therefore taken the interval $x - x_0$ too large. With five approximations we do as well here as we did with four above. The method is not taken to the extreme limits in Fig. 123.

In many cases, the representation of the first approximation is not as simple as it is here. If, starting from an arbitrary approximation, we have approximated only a small piece of the integral curve, then we can usually depend on our judgment, and extrapolate the portion from the first interval, finding the first approximation for the second interval. The greater the portion of the integral curve already plotted, the greater is the certainty of being able to extend this portion.

3. To apply the numerical method of stepwise approximation, we use one of the formulas derived in Art. 12 for the numerical integration. We select the Bessel formula most convenient for calculation in the fo[illegible]y derived from 12(6):

$$k_n = y_{\cdots}, \qquad y_n = h \cdot \int_{x_n}^{x_n+h} y'\,dx = h\left(\frac{y'_n + y'_{n+1}}{2} - \frac{1}{12}\overline{\Delta^2_{n+(1/2)}} \right. \tag{6}$$

$$\left. + \frac{11}{720}\overline{\Delta^4_{n+(1/2)}} - \frac{191}{60480}\overline{\Delta^6_{n+(1/2)}} \cdots \right).$$

We continue the difference scheme to the end of the interval in this way, while we take the last calculated difference as constant.

4. Values of y are recorded in the following scheme. These have been read from the last approximation curve for equidistant values of x. These values are denoted by $y_{(1)}$.

TABLE I.

x	$y_{(1)}$	$x \cdot y_{(1)}$	$y'_{(1)}$	Δ^1	Δ^2	k	$y_{(2)}$
1	2	2	−1		+155		2
			−1.0215	−430	*12*	−0.1027	
1.1	1.898	2.0878	−1.0430		+135		1.8973
			−1.0578	−295	*10*	−0.1088	
1.2	1.792	2.1504	−1.0725		+115		1.7985
			−1.0815	−180	*10*	−0.1025	
1.3	1.684	2.1892	−1.0905		+114		1.6860
			−1.0938	− 66	*9*	−0.1047	
1.4	1.574	2.2036	−1.0971		+108		1.5713
			−1.0950	+ 42	*8*	−0.1058	
1.5	1.463	2.1945	−1.0929		+ 87		1.4655
			−1.0865	+129	*8*	−0.1073	
1.6	1.354	2.1664	−1.0800		+ 97		1.3582
			−1.0687	+226	*7*	−0.1094	
1.7	1.246	2.1182	−1.0574		+ 74		1.2488
			−1.0424	+300	*7*	−0.1031	
1.8	1.142	2.0556	−1.0274		+ 95		1.1457
			−1.0077	+395	*7*	−0.1084	
1.9	1.040	1.9760	−0.9879		+ 61		1.0473
			−0.9651	+456	*7*	−0.0958	
2	0.944	1.8880	−0.9423		+ 95		0.9415
			−0.9148	+551	*7*	−0.0955	
2.1	0.851	1.7871	−0.8872		70		0.8560
			−0.8562	+621	*5*	−0.0867	
2.2	0.764	1.6808	−0.8251		+ 59		0.7693
			−0.7911	+680	*4*	−0.0715	
2.3	0.684	1.5732	−0.7571		+ 44		0.6978
			−0.7209	+724	*3*	−0.0712	
2.4	0.612	1.4688	−0.6847		+ 20		0.6166
			−0.6475	+744	*2*	−0.0677	
2.5	0.549	1.3725	−0.6103		+ 26		0.5589
			−0.5718	+770	*3*	0521	
2.6	0.494	1.2844	−0.5333		+ 44		0.4968
			−0.4926	+814	*2*	0.0428	
2.7	0.446	1.2042	−0.4519		+ [illegible]		0.4440
			−0.4109	+820	− [illegible]	0.0408	
2.8	0.406	1.1368	−0.3699		− 32		0.4032

Then the interval is $h = 0.1$. The numerical approximation method may now be applied to the entire interval from $x = 1$ to $x = 2.8$ (which is certainly too large) in order to illustrate the operation of the method (Table 1). Here we first calculate xy (column 3), subtract 1, and take the square root of the remainder. This gives the values of $y'_{(1)}$ entered in column 4. We form the difference scheme for these values in units of the last decimal place. This is carried only to the second difference which, because of the inaccuracy of the values read

off from the drawing, is very irregular. We then record the intermediate values (italics) and one twelfth of the intermediate values in the column Δ^2 and calculate the values k_1 by the Bessel formula (next to the last column). We can obtain the next approximation $y_{(2)}$ with these values k_1. The $y_{(2)}$ are entered in the last column. We observe that the difference between the initial value and the values of the second approximation becomes larger, in general, as we go further away from the initial point; nevertheless the differences remain smaller than 0.6 mm. with respect to the values read off from the drawing. In the second approximation, one more place is carried out than in the first. We may form a third approximation with these values in the same way, and proceed similarly until the latest values calculated no longer differ from the values of the previous approximation. In this case we do not record additional approximations for such values. The calculation is carried through to the fifth decimal, since we can neglect the differences of fourth order.

TABLE II

x	$y_{(1)}$	$y_{(2)}$	$y_{(3)}$	$y_{(4)}$	$y_{(5)}$	$y_{(6)}$	$y_{(7)}$	$y_{(8)}$	$y_{(9)}$	$y_{(10)}$
1	2	—								
1.1	1.898	1.8973	1.89774							
1.2	1.792	1.7985	1.79187							
1.3	1.684	1.6860	1.68634							
1.4	1.574	1.5713	1.57418							
1.5	1.463	1.4655	1.46454							
1.6	1.354	1.3582	1.35570							
1.7	1.246	1.2488	1.24857	1.24859						
1.8	1.142	1.1457	1.14403	1.14408						
1.9	1.040	1.0473	1.04290	1.04302	1.04301					
2.0	0.944	0.9415	0.94598	0.94620	0.94617					
2.1	0.851	0.8560	0.85401	0.85438	0.85432					
2.2	0.764	0.7693	0.76768	0.76828	0.76818					
2.3	0.684	0.6978	0.68762	0.68860	0.68838	0.68841				
2.4	0.612	0.6166	0.61440	0.61598	0.61555	0.61563	0.61562			
2.5	0.549	0.5589	0.54853	0.55103	0.55021	0.55041	0.55037			
2.6	0.494	0.4968	0.49049	0.49429	0.49277	0.49322	0.49311	0.49313		
2.7	0.446	0.4440	0.44068	0.44624	0.44345	0.44447	0.44417	0.44424	0.44423	
2.8	0.406	0.4032	0.39933	0.40725	0.40215	0.40445	0.40361	0.40399	0.40382	0.40

We see from the path of the approximating function that we have taken the interval too large. In particular, for $x = 2.8$, there is n convergence, since the value $y_{(4)}$, for example, is further distant fron the true value than was the initial value. We can therefore reduce the work of calculation if we proceed stepwise. First we improve only the first three values, then we correct the fourth until it no longer

changes, etc. The final values lie a little further than 0.4 mm. distant from those found by drawing. The work of computation would not have been much greater if we had started out from far worse initial values.

The advantage of the approximation method is that errors occasionally made in the course of the calculation are automatically corrected. In addition, large errors of calculation are quickly noticed because of the irregularities which appear in the representation of the difference table for y'. They should then be corrected at once, since otherwise they can make a manifold repetition of the process necessary.

5. In the case where the differences of fourth order no longer need to be considered (as in the case discussed above), we can avoid the construction of the difference scheme and use one of the quadrature formulas for integration given in Art. 15 and Art. 16. Because of the small effect of the third ordinate, the special Simpson formula given in 15(23) is most suitable. The equation (6) would then become:

$$y_{n+1} = y_n + \int_{x_n}^{x_n+h} y'\,dx = y_n + (5y'_n + 8y'_{n+1} - y'_{n+2})\frac{h}{12}. \tag{7}$$

In the example of the preceding section, we would perhaps have taken the values from 1 to 1.4, and calculated according to the following scheme:

x	$y_{(1)}$	$y'_{(1)}$	$y_{(2)}$	$y_{(3)}$	$y_{(4)}$	$y_{(5)}$
1	2	−1	2	2		
1.1	1.898	−1.0430	1.89774	1.89775		
1.2	1.792	−1.0725	1.79186	1.79189	1.79188	
1.3	1.684	−1.0905	1.68362	1.68366	1.68365	1.68365
1.4	1.574	−1.0971	1.574	1.574	1.57419	1.57418
1.5				1.463	1.463	1.46460
1.6					1.354	1.354
1.7						1.246

We would then have calculated $y'_{(1)}$ from the approximation values x, $y_{(1)}$ by means of the equation $y' = -(xy - 1)^{1/2}$. Then we would have calculated the approximation values $y_{(2)}$ up to $x = 1.3$, by means of the above formula:

$$y_{(2)}(1.1) = 2 + 0.1(-5 - 8.3440 + 1.0725)\cdot\frac{1}{12} = 1.89774. \tag{7b}$$

A repetition of the calculation gives only a negligible change for $y(1.1)$ and $y(1.2)$; we therefore do not consider any further calculation of this value, but consider the value of y for $x = 1.5$. This gives the

corrected value $y_{(4)}$. Here $y(1.2)$ remains unchanged; we calculate new approximations for $y(1.3)$ to $y(1.5)$ etc. We see that more rapid progress results with this stepwise procedure than would have been the case with the larger interval. This is also true with the use of the difference formula instead of Simpson's rule.[5]

NOTES

1. See Picard, *Traité d'Analyse,* 2nd ed. (Paris, 1905), Ch. XI, III.
2. Runge, *Jahresbericht der deutschen Math.-Ver.* **16** (1907), p. 270.
3. Runge-König, *Numerisches Rechnen* (Berlin, 1924), p. 300.
4. Groenveld, *Z. f. angew. Math. u. Mech.* **7** (1927), p. 150.
5. Lindow shows how other formulas derived in Art. 12 can be used for the integration of differential equations of first order by stepwise approximation. *Infinitesimalrechnung* (Berlin, 1928), Ch. IV, 1-5.

34. Approximate Integration of Differential Equations of Second and Higher Orders.

1. While the methods described in the last two paragraphs for the approximate determination of the integral curve can also be applied to the integration of equations of higher order, the extension of the *graphical methods* mentioned in Art. 31 has no practical importance. But we can easily get a general picture of the path of the integral curve of an equation of second order. The method is due to Lord Kelvin.[1] We draw *approximation curves fitted together from circular arcs.* These curves are so joined together that the tangent of the curve changes its direction continuously. As the radii of the individual circular arcs we use more or less accurate mean values of the radii of curvature of the pieces of the integral curve, as approximated by the circles under consideration.

We consider the equation

$$y'' = f(x, y, y')$$

and let $f(x, y, y')$ be a continuous function of its three arguments. This equation is now so transformed that the radius of curvature can be calculated for it as a function of the coordinates and the tangent angle. If $y' = \operatorname{tg} \tau$, then

$$y'' = \frac{1}{\cos^2 \tau} \frac{d\tau}{dx} = \frac{1}{\cos^3 \tau} \frac{d\tau}{ds} = \frac{1}{\cos^3 \tau} \frac{1}{\rho},$$

where ρ is to be taken as positive if it lies toward the forward direction as the y axis to the x axis. The general differential equation of second order can therefore be put in the form

$$\frac{1}{\rho} = \cos^3 \tau \cdot f(x, y, \operatorname{tg} \tau).$$

In case the right side of the equation is suitable, and if we want to trace out a rather large number of integral curves, we can sketch a nomogram before beginning the construction. From this we can evaluate ρ from the values for x, y and τ. Otherwise we must calculate ρ each time.

2. *In carrying out the construction* we can proceed in the following way, similar to that followed in 32.2: At the initial point A_1 we first calculate the value ρ_1 from the given values x_1, y_1, y_1'. We plot this value in the proper sense on the line perpendicular to the tangent direction, which is given by y_1'. We then obtain the point M_1, about which we describe a circular curve (of radius ρ_1) with the subtended angle α. For the endpoint B_1 of this arc we determine the coordinates x_2, y_2 and y_2', from which we now calculate ρ_2. We plot ρ_2 from C_1 out on the radius M_1C_1 belonging to half the central angle, $\alpha/2$. We then obtain the center point M_2 of the next approximation circle. We now describe a portion of the approximation circle about M_2 with ρ_2 as a radius. This curve has perhaps the central angle $3\alpha/2$. This arc cuts the radius M_1B_1 at the point A_2. If the point A_2 differs markedly from B_1, or if the slope of the new arc at A_2 is essentially different from that of the preceding one at B_1, then we calculate with the coordinates of A_2 and the corresponding slope again, and carry through the construction of the new arc. We repeat this until the new intersection point of the circular arc with M_1B_1 no longer differs from that of the preceding one, and until the slope is also the same as that previously determined. Now we consider the point of the curve A_2B_2 with the center M_2 and so forth. If we want to obtain a quick approximation, we can make the circle radius somewhat larger or smaller than is found from the calculation, according to its change from the very beginning. With some practice, a technique can be developed for estimating the increase or decrease of which is then necessary. We shall then very rapidly obtain a good general sketch of the path of the various integral curves. In each case it is advisable, for sketching the individual curves, to note the values of x, y and y' used for the calculation of the final values of the various ρ, since we can calculate the corresponding values y'' from these. Consequently we can correct the first sketch by the method of iteration, which is further described below.

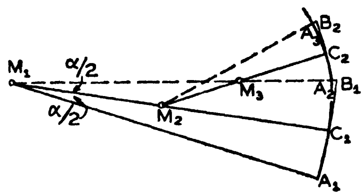


FIG. 124

3. A small *celluloid ruler*, as was first developed by *Boys*[2] and later revised by *Rothe*,[3] can facilitate the construction. The ruler has a hole S for the drawing pencil. Two mutually perpendicular lines intersect at this

point, one of which runs parallel to the ruler. Two marks, E_1 and E_2 are placed on the other line, each at a distance l from S, and serve for a convenient determination of y'. The line parallel to the edge has in general a uniform scale. The center M of a circle is placed on this scale, either by means of a pin inserted through the celluloid, or by means of the third leg of a small tripod, the other legs of which are set on the drawing paper. The angle subtended is indicated by special marks on the ruler. We can introduce still other scales on the straight edge, under certain conditions, in order to simplify the calculation of ρ. Most frequently a scale for $1/z$ will be useful. The determination of y' by means of the marks follows in this way: either we read the value on the scale of the straight edge of the point in which it is intersected by the x-parallel through E_2, which is ly'; or for very large values of y' we determine the value for the point at which it is cut by the y-parallel through E_1; this is l/y'.

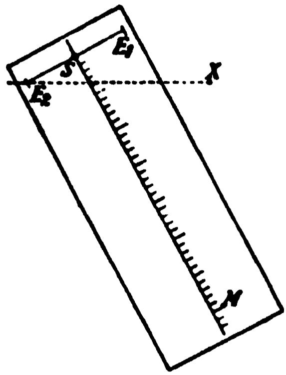


FIG. 125

If the curvature of the integral curve becomes very small at any place, (and therefore the radius of curvature becomes so large that it is impractical to approximate by arcs), then it is best to do the approximating by segments of straight lines, the change in slope of which is calculated from the equation $\Delta y' = y''\Delta x = f(x, y, y')\Delta x$. We can calculate y' itself in the neighborhood of points for which $y'' = 0$ from the equation $f(x, y, y') = 0$.

If we want to use the method with polar coordinates r, φ, we use the equation in the form

$$\frac{d^2r}{d\varphi^2} = r'' = f\left(r, \varphi, \frac{dr}{d\varphi}\right) = f(r, \varphi, r').$$

The advantage is that we can take r' as the polar subnormal directly from the figure; the disadvantage is that the calculation of

$$\rho_m = \frac{(r_m^2 + r_m'^2)^{3/2}}{r_m^2 + 2r_m'^2 - r_m \cdot r_m''}$$

is quite detailed,[4] while, with the use of cartesian coordinates, the determination of ρ is simpler.[5]

4. This method becomes especially convenient if we do not need to determine y'. This is the case, for example, in the determination of the *meridian curve of the lying or hanging drop*, the equation of which is given by Adams in the form

$$\frac{1}{\rho} + \frac{\sin\theta}{x} = 2 \pm \beta y = \frac{1}{\rho} + \frac{1}{r},$$

where ρ and r are the principal radii of curvature of the surface, i.e., ρ is the radius of curvature of the meridian curve and r is the length of its normal to the axis of rotation. In this way the origin 0 lies at the apex of the drop and the y axis lies in the direction of the axis of rotation, the positive direction toward the interior of the drop. At the apex of the drop, i.e., at the origin, $\rho = r$, and therefore $\rho = r = x/\sin\theta = 1$. The positive sign in the equation holds for the lying drop, while the negative sign holds for the hanging drop. Fig. 126 shows a sketch for the case $\beta = 6$. In this case the meridian curve is always approximated by an arc drawn between two successive (solid line) radii, while the dotted radii serve for the measurement of the value of r. If we introduce the values calculated by Adams[6] by the method of the calculus of differences, or by Koch[7] by the Runge-Kutta method in the drawing, then the very small deviations lie within the accuracy of the drawing.

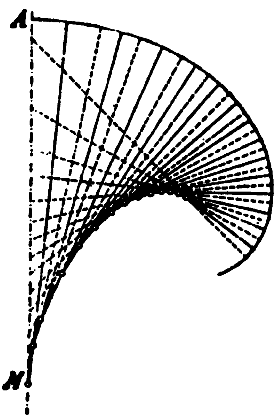


FIG. 126

5. The *numerical methods of integration* described in Art. 32 can be extended to systems of simultaneous linear differential equations of any order. Each such equation $y^{(n)} = f(x, y, y', \cdots, y^{(n-1)})$ can be transformed by introducing a new variable for each derivative. In a particularly simple system of simultaneous equations, we have

$$(1) \qquad \begin{aligned} &y' = u, \qquad y'' = u' = v \cdots, y^{(n-1)} = w' = z, \\ &z' = f(x, y, u, v, \cdots, w, z). \end{aligned}$$

In many cases we shall use other substitutions, perhaps to get a system of equations which is more convenient for numerical calculation. The derivations given in Art. 32 can also be extended to systems of linear equations. In these latter cases, the expressions do not become much more complicated, provided that we introduce correspondingly generalized expressions for the operators. For example, if we have two simultaneous equations

$$(2) \qquad y' = f(x, y, z), \qquad z' = g(x, y, z),$$

then we set

(2a) $$D'(\varphi) = \varphi_x + f\cdot\varphi_y + g\cdot\varphi_z\,,$$

etc. We can then obtain identity of terms with the factor h^4 in a way completely analogous to that of Art. 32. Here we give only the extension of the *formula* (3) *of Kutta* given in Art. 32, to two simultaneous equations:[8]

(3)
$$\begin{aligned}
k' &= f(x_0, \quad y_0, \quad z_0)h;\\
l' &= g(x_0, \quad y_0, \quad z_0)h\\
k'' &= f\left(x_0 + \frac{h}{2}, y_0 + \frac{k'}{2}, z_0 + \frac{l'}{2}\right)h;\\
l'' &= g\left(x_0 + \frac{h}{2}, y_0 + \frac{k'}{2}, z_0 + \frac{l'}{2}\right)h\\
k''' &= f\left(x_0 + \frac{h}{2}, y_0 + \frac{k''}{2}, z_0 + \frac{l''}{2}\right)h;\\
l''' &= g\left(x_0 + \frac{h}{2}, y_0 + \frac{k''}{2}, z_0 + \frac{l''}{2}\right)h\\
k^{(4)} &= f(x_0 + h, y_0 + k''', z_0 + l''')h;\\
l^{(4)} &= g(x_0 + h, y_0 + k''', z_0 + l''')h\\
k &= \frac{1}{6}(k' + 2k'' + 2k''' + k^{(4)})\\
l &= \frac{1}{6}(l' + 2l'' + 2l''' + l^{(4)}),
\end{aligned}$$

where k and l are the increments of y and z, which correspond to an increment h of x. Here also the error in the first approximation is proportional to h^5. We again obtain a measure for the resultant error if we carry out the calculation a second time with the doubled interval width and take perhaps 1/15 of the difference of the two results.

6. If, instead of a general system of linear equations, we have the simpler system which arises from an equation of higher order, e.g.,

(4) $$z' = g(x, y, z), \qquad y' = z,$$

then the equations (3) are materially simplified, since the increments k and l are obtained from the values for h. The system then becomes

$$
\begin{aligned}
k' &= z_0 \cdot h; & l' &= g(x_0, \quad y_0, \quad z_0 \quad)h \\
k'' &= \left(z_0 + \frac{l'}{2}\right)h; & l'' &= g\left(x_0 + \frac{h}{2}, y_0 + \frac{k'}{2}, z_0 + \frac{l'}{2}\right)h \\
k''' &= \left(z_0 + \frac{l''}{2}\right)h; & l''' &= g\left(x_0 + \frac{h}{2}, y_0 + \frac{k''}{2}, z_0 + \frac{l''}{2}\right)h \\
k^{(4)} &= (z_0 + l''')h; & l^{(4)} &= g(x_0 + h, y_0 + k''', z_0 + l''')h \\
k &= \frac{1}{6}(k' + 2k'' + 2k''' + k^{(4)}); & l &= \frac{1}{6}(l' + 2l'' + 2l''' + l^{(4)}).
\end{aligned} \tag{5}
$$

Nyström[9] has devised new formulas for a system of differential equations of second order which can be reduced to a system of equations of the form (4). These formulas yield the same approximation as the above with somewhat less calculation. However, we employ the above method because of its clear construction.

7. As an *example*, we take a *pendulum oscillation with a linear damping term*

$$\varphi'' + \beta\varphi' + \alpha \sin \varphi = 0, \tag{6}$$

where φ is the angular displacement, and $\beta = 0.1560$, $\alpha = 4.905$. The corresponding system of linear equations is

$$z' = -(\beta z + \alpha\varphi), \qquad \varphi' = z. \tag{7}$$

We take $\varphi = \pi/2$, $z = 0$ as the initial conditions, so that the pendulum swings down from the horizontal position without an initial impulse. The magnitude of the interval for t is taken as $h = 0.2$. Calculations are carried out according to the scheme on page 400 by use of the formula given above of Runge-Kutta. Both of these values are entered in the table.

The tables of Hayashi[10] are used in the calculation. These permit us to set up the trigonometric functions for the angle in radians, and consequently make a reduction to degrees unnecessary. The calculation is carried through up to $t = 5$ sec. The following values are obtained:

t	z	$k = h \cdot z$	$[k]$	φ	$\sin \varphi$	$-0.1560z$	$-4.905 \sin \varphi$	t	$l = h \cdot z$	$[l]$	z	φ
	0	0	0	1.5708	1	0	−4.9050	−4.9050	−0.9810	−0.9810	0	1.5708
	−0.4905	−0.0981	−0.1962	1.5708	1	+0.0765	−4.9050	−4.8285	−0.9657	−1.9314		
	−0.4829	−0.0966	−0.1932	1.5217	0.99880	+0.0753	−4.8991	−4.8238	−0.9648	−1.9296		
	−0.9648	−0.1930	−0.1930	1.4742	0.99534	+0.1505	−4.8821	−4.7316	−0.9463	−0.9463	−0.9647	−0.0971
0.2			−0.5824							−5.7883	−0.9647	+1.4737
	−0.9647	−0.1929	−0.1929	1.4737	0.99529	0.1505	−4.8819	−4.7314	−0.9463	−0.9463		
	−1.4378	−0.2876	−0.5751	1.3772	0.98132	0.2243	−4.8134	−4.5891	−0.9178	−1.8356		
	−1.4236	−0.2847	−0.5694	1.3299	0.97113	0.2221	−4.7634	−4.5413	−0.9083	−1.8165		
	−1.8730	−0.3746	−0.3746	1.1890	0.92800	0.2922	−4.5518	−4.2596	−0.8519	−0.8519	−0.9084	−0.2853
0.4			−1.7120							−5.4503	−1.8731	+1.1884

t	z	φ	t	z	φ
0	0	1.5708	2.6	+2.4448	+0.3534
0.2	−0.9647	1.4737	2.8	+1.8430	+0.7877
0.4	−1.8731	1.1884	3.0	+1.0057	+1.0749
0.6	−2.6072	0.7358	3.2	+0.0956	+1.1853
0.8	−2.9487	0.1720	3.4	−0.7942	+1.1145
1.0	−2.7420	−0.4061	3.6	−1.5840	+0.8742
1.2	−2.0696	−0.8931	3.8	−2.1490	+0.4959
1.4	−1.1618	−1.2184	4.0	−2.3383	+0.0400
1.6	−0.1957	−1.3543	4.2	−2.0855	−0.4093
1.8	+0.7506	−1.2981	4.4	−1.4783	−0.7702
2.0	+1.6227	−1.0588	4.6	−0.6821	−0.9881
2.2	+2.3064	−0.6614	4.8	+0.1657	−1.0398
2.4	+2.6212	−0.1612	5.0	+0.9703	−0.9248

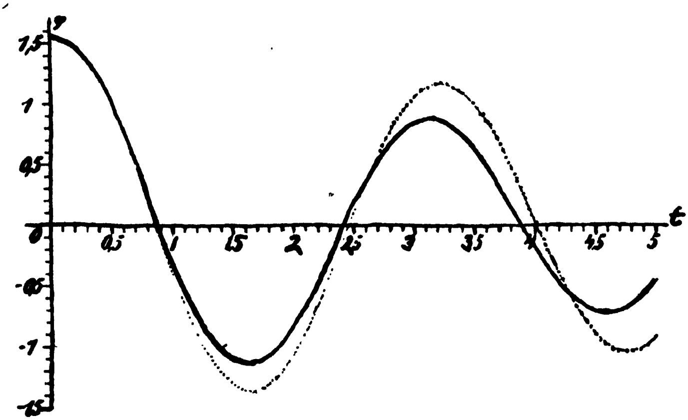


FIG. 127

The calculated values are plotted in Fig. 127, and are connected by the dotted curve. The curve which we get as the integral curve of the above equation is plotted for comparison. Here we replace φ' by $(\varphi')^2$, i.e., we introduce a quadratic damping term. This determination of the curve is performed graphically in Fig. 128, and the curve is given here only for comparison purposes.

8. The *second method* considered in 32(8) for integration can also be extended to systems of equations and to equations of higher order. For example, if we have two simultaneous linear equations

$$y' = f(x, y, z), \qquad z' = g(x, y, z),$$

then we construct a difference scheme for f and g up to y'_m and z'_m. We

calculate new values y_{m+1} and z_{m+1} from these, by means of the formula given in 32(7). By substitution of these values in the functions f and g,

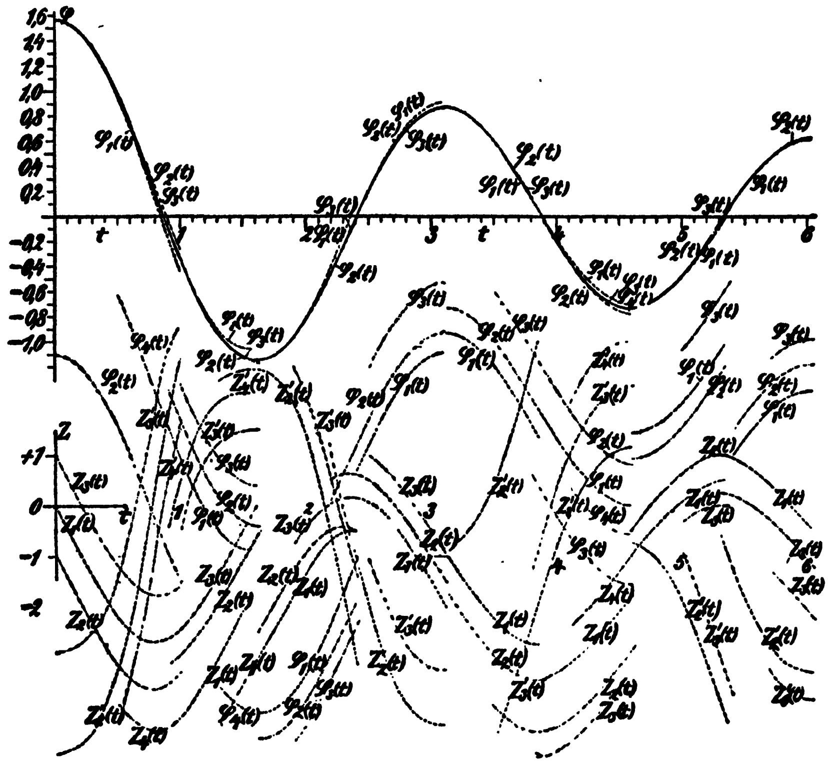


FIG. 128

we obtain the values y'_{m+1} and z'_{m+1} with which the difference scheme is continued.

The calculation is much simpler with an ordinary differential equation of higher order. For example, if we have an equation of second order

$$y'' = f(x, y, y'), \tag{8}$$

then we need only prepare a difference scheme for y''. From this scheme we calculate the next value of y', by means of the formula

$$\begin{aligned} y'_{m+1} = y'_m + h\Big(y''_m + \frac{1}{2}\,\Delta^1_{m-(1/2)} + \frac{5}{12}\,\Delta^2_{m-1} + \frac{3}{8}\,\Delta^3_{m-(3/2)} + \frac{251}{720}\,\Delta^4_{m-2} \\ + \frac{95}{280}\,\Delta^5_{m-(5/2)}\,\cdots\Big) + R_{m+1}\,, \end{aligned} \tag{9}$$

which corresponds to the formula of 32(7), and which follows from 12(1). We get the next value of y_{m+1} from the formula following from 12(18):

$$(10)\quad \begin{aligned} y_{m+1} &= 2y_m - y_{m-1} + h^2 y_m'' \\ &\quad + \frac{h^2}{12}\Big(\Delta^2_{m-1} + \Delta^3_{m-(3/2)} + \Delta^4_{m-2} + \Delta^5_{m-(5/2)} - \cdots \\ &\quad - \frac{1}{20}\Delta^4_{m-2} - \frac{1}{10}\Delta^5_{m-(5/2)} \cdots\Big) + R_m + \overline{R}_m\,. \end{aligned}$$

By substitution of both these values in the function f we get the value y_{m+1}'', with which we continue the difference scheme. We can calculate the beginning of the scheme either by series development or by the Runge-Kutta formulas, or by the method of iteration to be developed below.

9. As an *example*, we take the equation which was treated above by means of the Runge-Kutta formula:

$$(6)\qquad \varphi'' = -\beta\varphi' - \alpha\sin\varphi$$

with the initial conditions $\varphi_0 = \pi/2$, $\varphi_0' = 0$. From this it follows that $\varphi_0'' = -\alpha$. To be able to calculate the first values of the difference scheme from a series development, we differentiate the above equation repeatedly. This gives

$$\varphi''' = -\beta\varphi'' - \alpha\varphi'\cos\varphi \qquad \varphi_0''' = +\beta\alpha$$

$$\varphi^{(4)} = -\beta\varphi''' - \alpha\varphi''\cos\varphi + \alpha\varphi'^2\sin\varphi \qquad \varphi_0^{(4)} = -\beta^2\alpha$$

$$\begin{aligned}\varphi^{(5)} &= -\beta\varphi^{(4)} - \alpha\varphi'''\cos\varphi + 3\alpha\varphi'\varphi''\sin\varphi \\ &\quad + \alpha\varphi'^3\cos\varphi \qquad \varphi_0^{(5)} = +\beta^3\alpha\end{aligned}$$

$$\begin{aligned}\varphi^{(6)} &= -\beta\varphi^{(5)} - \alpha\varphi^{(4)}\cos\varphi + 4\alpha\varphi'''\varphi'\sin\varphi + 3\alpha\varphi''^2\sin\varphi \\ &\quad + 6\alpha\varphi'^2\varphi''\cos\varphi - \alpha\varphi'^4\sin\varphi \qquad \varphi_0^{(6)} = -\beta^4\alpha + 3\alpha^3\end{aligned}$$

$$\begin{aligned}\varphi^{(7)} &= -\beta\varphi^{(6)} - \alpha\varphi^{(5)}\cos\varphi + 5\alpha\varphi^{(4)}\varphi'\sin\varphi + 10\alpha\varphi''\varphi'''\sin\varphi \\ &\quad + 10\alpha\varphi'''\varphi'^2\cos\varphi - 15\alpha\varphi''^2\varphi'\cos\varphi - 10\alpha\varphi'^3\varphi''\sin\varphi \\ &\quad - \alpha\varphi'^5\cos\varphi \qquad \varphi_0^{(7)} = +\beta^5\alpha - 13\beta\alpha^3\end{aligned}$$

$$\varphi^{(8)} = -\beta\varphi^{(7)} - \alpha\varphi^{(6)}\cos\varphi + 6\alpha\varphi^{(5)}\varphi'\sin\varphi + 15\alpha\varphi^{(4)}\varphi''\sin\varphi$$

$$+ 15\alpha\varphi^{(4)}\varphi'^2 \cos\varphi + 10\alpha\varphi'''^2 \sin\varphi + 60\alpha\varphi'\varphi''\varphi''' \cos\varphi$$

$$- 20\alpha\varphi'''\varphi'^3 \sin\varphi + 15\alpha\varphi''^3 \cos\varphi - 45\alpha\varphi''^2\varphi'^2 \sin\varphi$$

$$- 15\alpha\varphi'^4\varphi'' \cos\varphi + \alpha\varphi'^6 \sin\varphi \qquad \varphi_0^{(8)} = -\beta^6\alpha + 38\beta^2\alpha^3.$$

If we calculate these values for the given β and α and substitute in the series expansion of φ (in powers of t), we get

$$\varphi = 1.570796 - 2.45250t^2 + 0.12753t^3 - 0.0049737t^4$$

$$+ 0.00015519t^5 + 0.49171t^6 - 0.047485t^7 + 0.0027066t^8 + \cdots$$

and from this, by differentiation,

$$\varphi' = - 4.90500t + 0.38259t^2 - 0.019895t^3 + 0.0007759t^4$$

$$+ 2.9502t^5 - 0.33239t^6 + 0.021653t^7 + \cdots .$$

By substitution of the values of t in these two series, we get the first five values of φ and φ', which are entered in the second and third columns of the difference scheme below, down to the horizontal line. If we substitute these values in the given equation, then we get the corresponding values of φ'' which are entered in the fourth column of the scheme, also to the horizontal line. The difference scheme is formed from these values.

In order to consider how many places can be kept if we break off the quadrature formulas after the fourth difference, we estimate the remainder term. The remainder term of the first interpolation series is

$$| R_5 | = h^6 f^{(5)}(\tau) \int_0^1 \frac{t(t+1)(t+2)(t+3)(t+4)}{5!}\, dt.$$

Since the scheme is set up for φ'', then we have (32.9):

$$| R_5 | \approx h^6\varphi^{(7)}(\tau) \frac{40}{120} \approx \frac{1}{3} \cdot 240h^6 \approx 80h^6,$$

where, for $\varphi^{(7)}$, the value is taken at the point $t = 0$. In the scheme, $h = 0.05$, so that

$$| R_5 | \approx 1 \times 10^{-6}.$$

In the first step therefore, an error already exists in the sixth decimal. We can therefore carry out the calculation to at most five places. A much smaller error, namely 5×10^{-9}, is obtained if we break off the second integration formula after the fourth difference. We shall there-

fore carry through five places, under the assumption that the numerical values given in the differential equation are precise. Otherwise we must be satisfied with fewer places.

The difference scheme now becomes

t	φ	φ'	φ''	Δ^1	Δ^2	Δ^3	Δ^4
0.00	1.57080	0	−4.90500				
				+ 3821			
0.05	1.56468	−0.24430	−4.86679		+ 97		
				+ 3918		+ 323	
0.10	1.54640	−0.48667	−4.82761		+ 420		+205
				+ 4338		+ 528	
0.15	1.51605	−0.72699	−4.78423		+ 948		+189
				+ 5286		+ 717	
0.20	1.47374	−0.96493	−4.73137		+1665		+174
				+ 6951		+ 891	
0.25	1.41961	−1.19984	−4.66186		+2556		+124
				+ 9507		+1015	
0.30	1.35383	−1.43068	−4.56679		+3571		
				+13078			
0.35	1.27664	−1.65591	−4.43601				
0.40	1.18837	−1.87349					
...	...	...	...				

From the values above the horizontal lines we next form

$$\varphi_5' = \varphi_4' + h\left(\varphi_4'' + \frac{1}{2}\,\Delta^1_{7/2} + \frac{5}{12}\,\Delta^2_3 + \frac{3}{8}\,\Delta^3_{5/2} + \frac{251}{720}\,\Delta^4_2\right)$$

$$= -0.96493 + 0.05(-4.73137 + 0.02643 + 0.00395 + 0.00198 + 0.00072)$$

$$= 1.19984.$$

In exactly the same way, we calculate

$$\varphi_5 = 2\varphi_4 - \varphi_3 + h^2\varphi_4'' + \frac{h^2}{12}\left(\Delta^2_3 + \Delta^3_{5/2} + \Delta^4_2 - \frac{1}{20}\,\Delta^4_2\right)$$

$$= 2.94748 - 1.51605 + 0.0025\left[-4.73137 + \frac{1}{12}\,(0.00948 + 0.00528 + 0.00195)\right]$$

$$= 1.41961.$$

We calculate φ'' with these values from the given equation, and add this in the fourth column of the scheme. We then calculate the further differences from which we then obtain φ_6', φ_6 and φ_6'', etc.

A comparison with the values of Sec. 8 shows good agreement in the values of φ, even if small deviations in the values of φ appear, which increase with increasing t. The values found here are the more precise because of the smaller interval.

10. Finally, we must mention the extension of the method of iteration to systems of linear equations, and consequently to differential equations of higher order. We shall illustrate the method with two simultaneous equations

$$y' = f(x, y, z), \qquad z' = g(x, y, z), \tag{11}$$

which we immediately write in the form of integral equations

$$y = y_0 + \int_{x_0}^{x} f(x, y, z)\,dx, \qquad z = z_0 + \int_{x_0}^{x} g(x, y, z)\,dx. \tag{12}$$

If we now have any approximate solutions $y_{(1)}(x)$ and $z_{(1)}(x)$ of the above equations, we substitute these in the functions under the integral signs, and carry out the integrations. This gives us new approximations $y_{(2)}(x)$ and $z_{(2)}(x)$. We continue with these in exactly the same way, etc.; for example, we get the $(n + 1)$st approximations

$$\begin{aligned} y_{(n+1)} &= y_0 + \int_{x_0}^{x} f(x, y_{(n)}, z_{(n)})\,dx, \\ z_{(n+1)} &= z_0 + \int_{x_0}^{x} g(x, y_{(n)}, z_{(n)})\,dx. \end{aligned} \tag{13}$$

To find the *error of this approximation,* we subtract (13) from (12):

$$\begin{aligned} y - y_{(n+1)} &= \int_{x_0}^{x} (f(x, y, z) - f(x, y_{(n)}, z_{(n)}))\,dx; \\ z - z_{(n+1)} &= \int_{x_0}^{x} (g(x, y, z) - g(x, y_{(n)}, z_{(n)}))\,dx. \end{aligned} \tag{14}$$

The integrand can be transformed in the following way:

$$\begin{aligned} y - y_{(n+1)} = \int_{x_0}^{x} \Bigg[&\frac{f(x, y, z) - f(x, y_{(n)}, z)}{y - y_{(n)}} (y - y_{(n)}) \\ &+ \frac{f(x, y_{(n)}, z) - f(x, y_{(n)}, z_{(n)})}{z - z_{(n)}} (z - z_{(n)}) \Bigg] dx, \end{aligned} \tag{15}$$

$$z - z_{(n+1)} = \int_{x_0}^{x} \left[\frac{g(x, y, z) - g(x, y_{(n)}, z)}{(y - y_{(n)})} (y - y_{(n)}) \right.$$

$$\left. + \frac{g(x, y_{(n)}, z) - g(x, y_{(n)}, z_{(n)})}{z - z_{(n)}} (z - z_{(n)}) \right] dx.$$

The difference quotients appearing in the integrands are the mean values of the first partial derivatives of the functions f and g for values of the variables which lie in the intervals from y to $y_{(n)}$ and z to $z_{(n)}$. If we now assume that the absolute values of both derivatives of f in the interval under consideration are smaller than M_f, and those of g smaller than M_g, then, if we denote the maximum absolute value of the differences $y - y_{(n)}$ and $z - z_{(n)}$ in the interval from x_0 to x by $|\delta_{(n)}|$ and $|\epsilon_{(n)}|$, we get (by the mean value theorem)

$$\text{(15a)} \qquad \begin{aligned} |\delta_{(n+1)}| &\leqq M_f(|\delta_{(n)}| + |\epsilon_{(n)}|) \cdot |x - x_0|; \\ |\epsilon_{(n+1)}| &\leqq M_g(|\delta_{(n)}| + |\epsilon_{(n)}|) \cdot |x - x_0|, \end{aligned}$$

i.e.,

$$\text{(16)} \qquad |\delta_{(n+1)}| + |\epsilon_{(n+1)}| \leqq (M_f + M_g)(|\delta_{(n)}| + |\epsilon_{(n)}|) \cdot |x - x_0|.$$

If we choose the integration interval so small that

$$\text{(16a)} \qquad |M_f + M_g| \cdot |x - x_0| = K < 1,$$

then

$$\text{(16b)} \qquad |\delta_{(n+1)}| + |\epsilon_{(n+1)}| \leqq K(|\delta_{(n)}| + |\epsilon_{(n)}|) \leqq K^n(|\delta_{(1)}| + \epsilon_{(1)}|),$$

i.e., the deviation of the approximation value from the true value becomes arbitrarily small by continuation of the process, provided that we do not make the interval too large. The actual method is the following: we start from the given initial values x_0, y_0, z_0, and seek first approximations $y_{(1)}$, $z_{(1)}$ in a small interval. These approximations can be very rough. We improve these values by iteration. In this case we always use the best approximation values. For example, if we have determined $y_2(x)$, then we substitute the value $z_1(x)$ and $y_2(x)$ in g, in order to find $z_2(x)$ by integration of this function. How large we ought to choose the interval is determined from the convergence of the individual successive approximations themselves. If the approximation is carried through in this first interval until the latest approximation no longer differs from the preceding one, then we continue this function in another interval, in approximate fashion, apply the approximation method just described, etc. In this way we can approximate the solution piecewise.

We can exchange variables in systems of equations, just as in the case

of linear equations, if the function values become too large. For example, if we want to take y as the independent variable, then we choose the system

$$\text{(16c)} \qquad \frac{dx}{dy} = \frac{1}{f(x, y, z)}; \qquad \frac{dz}{dy} = \frac{g(x, y, z)}{f(x, y, z)}$$

instead of (11).

11. *Example*: For the graphical construction, we take the equation of the damped pendulum vibration already discussed above, except that we set the damping term proportional to the square of the velocity,

$$\text{(16d)} \qquad \varphi'' = \mp\beta\varphi'^2 - \alpha \sin \varphi,$$

where the sign of the first term on the right side must always be different from the sign of φ'. By the substitution $z = \varphi'$ we get the simultaneous equations

$$\text{(16e)} \qquad z' = \mp\beta z^2 - \alpha \sin \varphi; \qquad \varphi' = z.$$

For the constants we choose the same values as above, $\beta = 0.156$, $\alpha = 4.905$. Fig. 128 shows the method of the construction. It is best to start from a first assumption $z_{(1)}$ of the velocity curve (---- in the figure). We perhaps know this path approximately, since we know something about the operation of the process. We integrate this and get the curve $\varphi_{(1)}(t)$. With the z and φ values of this curve, we calculate individual values z' by substitution in the given equation. We plot these and draw a smooth curve $z_{(2)}(t)$ through them. We integrate this graphically and get the curve (-·-·-). If we integrate this again, we get the curve $\varphi_{(2)}(t)$ (also -·-·-). With $\varphi_{(2)}(t)$ and $z_{(2)}(t)$, we calculate the ordinates of the curve $z'_{(3)}(t)$, with which we proceed in the same way as with $z'_{(2)}$, etc. The curves $\varphi(t)$ are drawn next to each other in Fig. 128. The last value $\varphi_{(n)}$, which no longer differs from the previous, is drawn in. The starting points of the curves $\varphi(t)$, $z(t)$ and $z'(t)$ are shifted somewhat in the direction of the ordinate axis for greater clarity, and the starting points of the first curves $z(t)$ and $z'(t)$ are marked by coordinate crosses. If we have the final curves in the first interval, which goes from 0 to 1 in this case, we continue the z curve, for whose continuation the path in the first interval gives an approximation, into the second interval, and proceed as in the first interval. The final curve is extended according to Fig. 127 in order to make possible a comparison with the curve obtained with a linear damping term.

12. For the numerical solution of the differential equations of second order by successive approximation, we can use the formulas (9) and (10) of this article, or one of the other formulas of Art. 12, especially (6) and (20).

Example: We consider the capillarity equation discussed in Sec. 4 of this article. This may be written in the form of two simultaneous equations

$$\text{(16f)} \qquad \frac{dx}{d\varphi} = \frac{\cos\varphi}{2 + \beta y - (\sin\varphi)/x}, \qquad \frac{dy}{d\varphi} = \frac{\sin\varphi}{2 + \beta y - (\sin\varphi)/x}$$

where the slope angle φ of the tangent appears as independent variable. Again let $\beta = 6$. Further the condition $(\sin\varphi)/x = 1$ holds at the origin, so that $\varphi = 0$. The first approximation values are taken from the curve. As it turns out, the error of this value is, in general, not larger than 0.2 mm., so that the value lies almost within the limit of accuracy of the reading. The calculation is again performed over an interval too large for practical use, in order to show how the solution is approximated. It is first carried out with the slide rule, and the integration formula 12(6)

$$\text{(16g)} \qquad y = h\left(\frac{y_0 + y_1}{2} - \frac{1}{12}\bar{\Delta}^2_{1/2} + \frac{11}{720}\bar{\Delta}^4_{1/2} \cdots\right)$$

is used. For an interval $h = 2.5° = 0.043673$, the fourth differences need no longer be considered. The values of the second differences are completely irregular because of the error of reading, as is to be seen from the first interval reproduced below:

φ_0	x_0	y_0	x'	Δ^1	Δ^2	x_1	y'	Δ^1	Δ^2	y_1
0	0.000	0.000	1			0	0			0
			1.001	1	*−17*	*437*	*215*	430	*−23*	*95*
2.5	0.043	0.002	1.001		−17	0.0437	0.0430		− 1	0.00095
			993	−16	*−17*	*433*	*645*	429	*−24*	*282*
5	0.086	0.004	0.985		−16	0.0870	0.0859		−46	0.00377
			969	−32	*− 8*	*424*	*1051*	383	*−25*	*459*
7.5	0.128	0.010	0.953		+ 0	0.1294	0.1242		− 3	0.00836
			937	−32	*− 6*	*409*	*1432*	380	*−29*	*626*
10	0.170	0.015	0.921		−12	0.1703	0.1622		−54	0.01462
			899	−44	*− 1*	*392*	*1785*	326	*−34*	*781*
12.5	0.210	0.024	0.877		+12	0.2095	0.1948		−14	0.02243
			861	−33	*− 3*	*376*	*2104*	312	*−40*	*920*
15	0.247	0.032	0.844		−17	0.2471	0.2260		−66	0.03163
			819	−50	*− 2*	*357*	*2383*	246	*−34*	*1042*
17.5	0.283	0.044	0.794		+13	0.2828	0.2506		− 2	0.04205
			776	−37	*− 2*	*339*	*2628*	244	*−28*	*1148*
20	0.316	0.054	0.757			0.3166	0.2750			0.05353

The mean values are written in italics. For the next interval the behavior of the differences is essentially regular. The successive approximation values are entered in the following table. These are calculated with a slide rule up to the third approximation term, and from then on with five place logarithms. Furthermore, the fourth differences are also considered. We see that the approximation is obtained by terms which lie successively above and below the final value. We would therefore be able to shorten the process if we started out from the mean values of two successive differences. Occasionally deviations are reduced to the inaccuracies of rounding off. No convergence exists at the end of the interval.

φ	x_0	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
0	0.000	0.0000	0.0000	0.0000	0.00000	0.00000	0.00000	0.00000			
2.5	0.043	0.0437	0.0435	0.0436	0.04354	0.04357	0.04355	0.04356			
5	0.086	0.0870	0.0864	0.0868	0.08659	0.08670	0.08665	0.08668	0.08666	0.08667	
7.5	0.128	0.1294	0.1685	0.1292	0.12875	0.12899	0.12897	0.12893	0.12890	0.12891	
10	0.170	0.1703	0.1694	0.1704	0.16966	0.17006	0.16986	0.16996	0.16991	0.16993	
12.5	0.210	0.2095	0.2089	0.2099	0.20907	0.20962	0.20933	0.20947	0.20940	0.20943	0.20942
15	0.247	0.2471	0.2467	0.2478	0.24690	0.24743	0.24707	0.24725	0.24716	0.24720	0.24719
17.5	0.283	0.2828	0.2827	0.2838	0.28273	0.28336	0.28294	0.28316	0.28305	0.28310	0.28309
20	0.316	0.3167	0.3167	0.3178	0.31663	0.31734	0.31687	0.31712	0.31700	0.31705	0.31704

φ	y_0	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9
0	0.000	0.00000	0.00000	0.00000	0.000000					
2.5	0.002	0.00095	0.00096	0.00095	0.000951	0.000949				
5	0.004	0.00377	0.00379	0.00377	0.003777	0.003772	0.003773	0.003772		
7.5	0.010	0.00836	0.00843	0.00839	0.008406	0.008393	0.008398	0.008395	0.008396	
10	0.015	0.01462	0.01476	0.01468	0.014724	0.014697	0.014704	0.014703	0.014705	
12.5	0.024	0.02243	0.02264	0.02252	0.022589	0.022551	0.022564	0.022553	0.022558	0.022557
15	0.032	0.03163	0.03189	0.03174	0.031841	0.031767	0.031806	0.031786	0.031796	0.031790
17.5	0.044	0.04205	0.04237	0.04219	0.042314	0.042212	0.042271	0.042241	0.042256	0.042247
20	0.054	0.05353	0.05393	0.05369	0.053850	0.053717	0.053798	0.053756	0.053777	0.053765

Naturally the last digits should not be used because of the inaccuracies of rounding off. If we want additional decimal places, we must work with seven place logarithms, or with a calculating machine. To continue the scheme for rather large values φ, we first calculate further approximation values for larger values of φ according to the method explained in Sec. 8. These values can be improved by iteration wherever necessary. For the practical application it is important that we add only one or two new values, on which the method is applied, until these values no longer change, and then go to the next value, as is shown in 33.5. In this way a great deal of calculation can be avoided.

NOTES

1. Kelvin, *Phil Mag.* V, **34** (1892), p. 443.
2. Boys, *Phil. Mag.* **36** (1893), p. 75.
3. Rothe, *Z. f. Math. u. Phys.* **64** (1917), p. 90.
4. Neuendorff, *Z. f. angew. Math. u. Mech.* **2** (1922), p. 131.
5. Meissner, *Schweizer Bauzeitung* **62** (1913), Nos. 15, 16.
6. Bashforth and Adams, *An Attempt to Test the Theories of Capillary Action* (Cambridge, 1883).
7. Köch, *Über die praktische Anwendung der Runge-Kuttaschen Methode.* Dissertation. (Göttingen, 1909).
8. Emden, *Gaskugeln* (Leipzig, 1907), p. 92.
9. Nyström, *Acta societatis scientiarum fennicae* **50** (1925) No. 13.
10. Hayashi, *Fünfstellige Tafeln der Kreis-u. Hyperbelfunktionen* (Berlin, 1921).

35. Integrators.

1. Apparatus for the integration of arbitrary ordinary differential equations, such as were given by Kriloff for a group of equations, are extremely complicated. In addition, these devices must have a special tracing point for each function entering into the differential equation. Each such point would be led along by hand on the corresponding curve. Of course the pen can be moved by some mechanical means. But, since such a mechanism can travel only on certain curves, perhaps straight lines, circles, ellipses, etc., or by sliding in a groove on a few selected curves, then the apparatus is no longer useful for the solution of any differential equation, but only for such in which the corresponding functions appear. In practice it is then of importance to choose this guiding system so that the integrator can be used in as many ways as possible.

Most *integrators* employ wheels with milled edges. In the instruments of Pascal and Jacob, these run directly on the drawing paper, while the wheel on other integrators, for example in those of Knorr, runs on a cylinder which can be shifted in the direction of its main axis by a screw of variable pitch.

2. Starting out from a construction given by Potier,[1] E. Pascal[2] has developed a rather large number of different devices for the integration of special equations.

The *integrator of Potier* is simpler than that of Abdank (14.8). It consists of a framework LL' which can be moved parallel to the x axis by means of two wheels rr. The differential carriage W with the tracing point F (which is moved along the given curve $y = f(x)$) can be moved on the one y-parallel rail of this framework, while the integral carriage W_1, the motion of which is recorded by the point S, is moved along the other. The displacement is effected by a milled wheel which makes an angle with respect to the x axis determined by the connecting bar D of the differential

and integral carriages. This directrix rod D can be rotated about a pin M of the differential carriage, and may be displaced by means of the integral carriage so that the plane of the milled wheel (which is perpendicular to the drawing plane) is always parallel to the rod D or, in the *integrator*

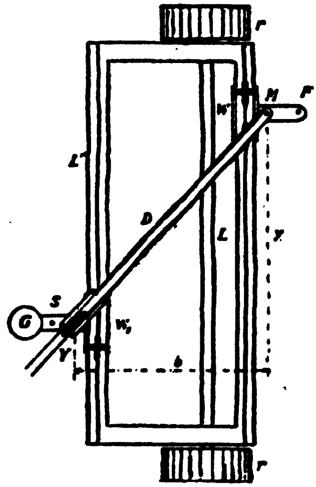


FIG. 129

of Pascal, can be set to a certain angle α = arc tg m with respect to this direction. For the first case we easily read from Fig. 129

$$Y' = \frac{f(x) - Y}{b}$$

where b is the x projection of the distance of the axis of rotation M from the contact point of the wheel. We can therefore integrate linear differential equations of first order

$$Y' + \frac{Y}{b} = \frac{1}{b} f(x)$$

with this instrument, where $f(x)$ is an arbitrary function. If the plane of the wheel forms the angle α = arc tg m with respect to the directrix, then the integrator draws the integral curve of the equation

$$\frac{Y' + m}{1 - mY'} = \frac{1}{b} [f(x) - Y].$$

If we add a second tracing point to the apparatus,[3] which permits us, in tracing on a curve $z = g(x)$, to make the separation of the two y-parallel

rails equal to $g(x)$, then the apparatus would, in the first case, integrate the general differential equation

$$Y' + \frac{1}{g(x)} Y = \frac{f(x)}{g(x)}.$$

A correspondingly generalized equation is obtained in the second case. The initial condition can be satisfied by an appropriate initial position of the integral carriage.

3. Since the simultaneous operation of two tracing points is always of doubtful accuracy, Pascal modified the instrument in another way. For the one group of apparatus, he replaced the straight edge by a curved rail which is capable of rotation either about the point M on the differential carriage, or about a point of the integral carriage, and which then can be shifted along the differential carriage. The plane of the integrating wheel is then parallel to the tangent to the curve, determined by the rail, at the point which has the same abscissa as the contact point of the wheel, or it forms a constant angle with respect to this direction. The Riccati and Abel differential equations, for example, can be integrated by such a device.

Pascal also replaced the straight rod L', on which the integral carriage runs, by a curved one. The equation of the path of a projectile in a resisting medium can be integrated with such an integrator. The resistance law determines the shape of the rail L'.

Pascal developed a second group of integrators from the device of Abdank-Abakanowitz (14.8). He made the pin P in Fig. 44 (through which the direction rail runs) movable. If this motion is independent of that of the integral carriage, then we can prepare an apparatus which integrates the equation of the path of the projectile for each resistance law which is given in graphical form. On the other hand, if the motion of P is dependent on that of the differential carriage, we can then solve integral equations.

Finally, a third group consists of the *polar integrators*, in which the two rails L and L' are no longer parallel, but form a constant angle with each other. This angle can be adjusted over an arc which is provided with a scale.

Other *integrators*, which do not permit us to draw the integral curve directly, but which do permit us to read off the value of the integral for each value of the independent variable, have been developed by *Jacob*[4] from the bar planimeter.

4. In conclusion, we shall describe the instrument of Knorr,[5] which uses the milled wheel R in connection with a cylinder T, movable along its axis (Fig. 130). The wheel rests on the outer surface of the cylinder. If the cylinder is rotated, then this is shifted on its axis so that the wheel

describes a helical path. The pitch of this helix is dependent on the angle which the plane of the wheel (perpendicular to the drawing plane) forms with the plane perpendicular to the axis of the cylinder. Since the axis of the cylinder is parallel to the plane of the drawing, this plane is also per-

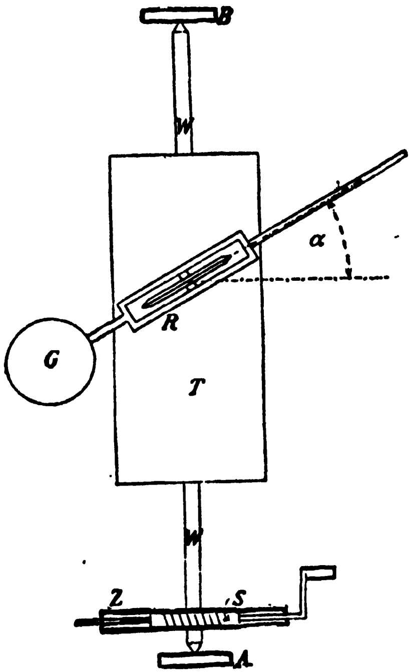


FIG. 130

pendicular to the drawing plane. Such screws of variable pitch were frequently used in earlier integrators.[6] In Knorr's apparatus, the rotation of the cylinder axis W is performed by means of a toothed wheel Z driven by the endless screw S. The cylinder drum T carries out the rotation of the axis because a pin N fits in a longitudinal groove of the shaft W (Fig. 131).

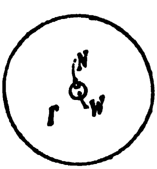


FIG. 131

The integrator of Knorr is used for the integration of equations of second order of the form

$$a \cdot y'' = f_1(y') + f_2(y) + f_3(x),$$

in which the three functions must be given as curves. For arbitrary initial conditions, the apparatus then draws the curves $y'(x)$ and $y(x)$. Since three arbitrary functions appear in the equation, the apparatus has three tracing pens, which are to be moved, by hand, over the corresponding curves: the point F_1 is moved

on the curve $f_1(y')$ drawn on the plane E_1, F_2 is moved on the curve $f_2(y')$ drawn on the plane E_2, and finally, F_3 is moved on the curve $f_3(x)$ drawn on the drum T_3. The two planes, E_1 and E_2, are each coupled with a screw of variable pitch, and are displaced, together with the drums T_1 and T_2, in the direction of their axes. In this way, the three tracing pens are always on the corresponding values y', y and x. The carriages with the three tracing pens slide along the rigid runners LL and their deflections are added algebraically by means of a cord (the broken line in Fig. 132) which is

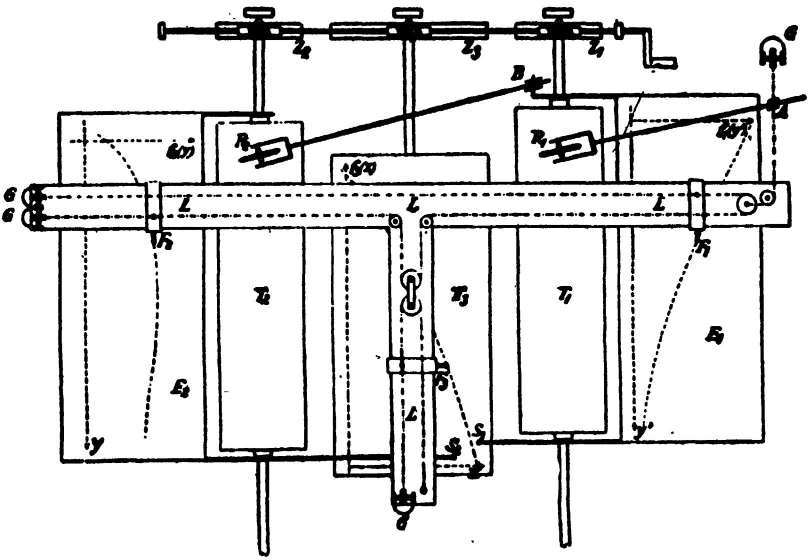


FIG. 132

stretched on these rails by means of weights. The values are transferred to the device A, which assigns an angle of inclination α to the wheel R_1, so that

$$a \cdot \operatorname{tg} \alpha = f_1(y') + f_2(y) + f_3(x).$$

Before the beginning of the integration, the entire apparatus is brought into the rest position. Then the calculation is carried out for the initial conditions, i.e., the tracing pen S_1, which draws on the drum T_3, and which is rigidly connected with the plane E_1 and the drum T_1, is moved (together with these) a distance $y'(x_0)$ from the x axis of the drum T_3. Also, the drawing pen S_2, which likewise writes on the drum T_3, is moved, together with the plane E_2 connected with it and the drum T_2, to a distance $y(x_0)$ from this axis. The tracing points F_1 and F_2 are then placed on the curves $f_1(y')$ and $f_2(y)$ of the drawing planes E_1 and E_2. If we now turn the drums through on an arc element dx, by means of the endless

screw and the toothed wheel X, while we also move the three tracing points on the diagrams, then the drum T_1 is turned in the direction of its axis by $[f_1(y') + f_2(y) + f_3(x)]\,dx/a = dy'/a$. The curve drawn by S_1 on the drum T_3 then has continuous tangents, the direction coefficients of which are proportional to $f_1(y') + f_2(y) + f_3(x)$ and whose ordinate is therefore proportional to $y'(x)$. The displacement of the plane E_1 now produces (by means of the direction control B) an inclination of the plane of the second wheel R_2 by an angle β, whose tangent will be proportional to $y'(x)$. The drawing pen S_2 will therefore draw a curve on the drum T_3 whose tangent has the value tg $\beta = \mu y'(x)$. Its ordinate will therefore be $\mu y(x)$, i.e., it will be the desired integral curve.

The integrator is used especially to construct the path, time, and velocity diagrams of a piston, if the resisting force $f_2(s)$ and the accelerating force $P = f_1(v)$ are given.

NOTES

1. Abdank-Abakanowitz, *Die Integraphen* (Leipzig, 1889), p. 166.
2. E. Pascal, *I miei integratori* (Naples, 1914). Abstracts from this are given by Galle, *Z. f. Instrumentenkunde* (1922).
3. Willers, *Z. f. Math. u. Phys.* 59 (1911), p. 36.
4. Jacob, *Mémorial de l'artillerie navale* (1909, 1910); *Le calcul mécanique* (Paris, 1911). Willers, *Mathematische Instrumente* (Berlin, 1926), Art. 19.
5. Knorr, *Organ für die Fortschritte des Eisenbahnwesens*, 79 (1924), p. 353.
6. Abdank-Abakanowitz, op. cit., p. 17 ff.

NAME INDEX

SUBJECT INDEX